

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2002, 21:30:16 ; Search time 1583.87 Seconds

(without alignments)
8842.964 Million cell updates/sec

Title: US-09-813-820-3

Perfect score: 849

Sequence: 1 GACGATTAAGTGAATAAT.....AACAGTAAAGATACCAAG 849

Scoring table: IDENTITY_NUC

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : GenEmbl:

1: gb.ba:*
2: gb.hg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vl:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.om:*
20: em.or:*
21: em.ov:*
22: em.pat:*
23: em.ph:*
24: em.pl:*
25: em.ro:*
26: em.sts:*
27: em.sy:*
28: em.un:*
29: em.vl:*
30: em.htgo.hum:*
31: em.htgo.in:*
32: em.htgo.ro:*
33: em.htgo.hum:*
34: em.htg.in:*
35: em.htg.ro:*
36: em.htg.other:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	849	100.0	3827	1	STACNA	M81736 Staphylococ
2	849	100.0	3827	6	AR067706	AR067706 Sequence
3	792.2	93.3	3468	6	A18434	A18434 Hybrid DNA
4	792.2	93.3	4612	6	A18436	A18436 assembled s
5	792.2	93.3	4612	6	AR067712	AR067712 Sequence
6	73.4	8.6	155204	2	AC007926	AC007926 Trypanoso
7	65.6	7.7	171798	2	AC084397	AC084397 Trypanoso
8	64.2	7.6	7218	6	I66494	I66494 Sequence 14
9	63.2	7.4	155204	2	AC007926	AC007926 Trypanoso
10	62.8	7.3	208546	2	AC083827	AC083827 Homo sapi
11	61.8	7.3	1822	1	AF260878	AF260878 Enterococ
12	61.8	7.3	318221	2	PFMAL13P3	AL049184 Plasmodu
13	61.4	7.2	104992	2	AC005504	AC005504 Plasmodu
14	61.4	7.2	169546	2	AC004157	AC004157 Plasmodu
15	60.6	7.1	110000	2	PFMAL4P1_1	Continuation (2 of
16	60.2	7.1	957	1	AF260880	AF260880 Enterococ
17	60.2	7.1	957	1	AF260881	AF260881 Enterococ
18	60.2	7.1	957	1	AF260882	AF260882 Enterococ
19	60.2	7.1	957	1	AF260883	AF260883 Enterococ
20	60.2	7.1	957	1	AF260888	AF260888 Enterococ
21	60.2	7.1	957	1	AF260890	AF260890 Enterococ
22	60.2	7.1	957	1	AF260893	AF260893 Enterococ
23	60.2	7.1	957	1	AF260895	AF260895 Enterococ
24	60.2	7.1	1963	1	AF260876	AF260876 Enterococ
25	58.8	6.9	192929	2	AC005505	AC005505 Plasmodu
26	58.8	6.9	256172	2	AC005139	AC005139 Plasmodu
27	58.6	6.9	863	11	CNS06EVO	AL395628 T7 end of
28	58.6	6.9	957	1	AF260889	AF260889 Enterococ
29	58.6	6.9	957	1	AF260894	AF260894 Enterococ
30	58.6	6.9	1963	1	AF260877	AF260877 Enterococ
31	58.6	6.9	2104	1	AF260874	AF260874 Enterococ
32	58.6	6.9	2104	1	AF260875	AF260875 Enterococ
33	58.6	6.9	2245	1	AF260872	AF260872 Enterococ
34	58.6	6.9	2245	1	AF260873	AF260873 Enterococ
35	58	6.8	3989	14	EPVSPHER	X77052 Entomopoxvi
36	57.6	6.8	110000	2	PFMAL4P1_0	AL034537 Plasmodu
37	57.4	6.8	4096	3	AF362374	AF362374 Dictyoste
38	57.4	6.8	47577	3	AF396436	AF396436 Tetrahyme
39	57.4	6.8	189893	2	AC013254	AC013254 Drosophyl
40	57	6.7	957	1	AF260884	AF260884 Enterococ
41	57	6.7	957	1	AF260885	AF260885 Enterococ
42	57	6.7	957	1	AF260886	AF260886 Enterococ
43	57	6.7	957	1	AF260892	AF260892 Enterococ
44	57	6.7	957	1	AF260896	AF260896 Enterococ
45	57	6.7	1374	1	AF159247	AF159247 Enterococ

ALIGNMENTS

RESULT 1
STACNA 3827 bp DNA BCT 17-AUG-1994
Staphylococcus aureus collagen adhesin (cna) gene, complete cds.
M81736
M81736.1 GI:387879
cna gene: collagen adhesin.
KEYWORDS
SOURCE
ORGANISM
Staphylococcus aureus
Staphylococcus aureus
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE
Pattl, J.M., Jonsson, H., Guss, B., Switalski, L.M., Wildberg, K.,
Lindberg, M. and Hook, M.
Molecular characterization and expression of a gene encoding a
Staphylococcus aureus collagen adhesin [published erratum appears
in J Biol Chem 1994 Apr 15;269(15):11672]
JOURNAL
MEDLINE
92165839
COMMENT
On Aug 30, 1993 this sequence version replaced gi:386712.

FEATURES

```

URES      location/Qualifiers
source     1..3827
           /organism="Staphylococcus aureus"
           /db_xref="taxon:1280"
           /lsuse_lib="FDA 574; lambda gt11
RBS        137..144
           /gene="cna"
gene        137..3702
           /gene="cna"
CDS         151..3702

```

BASE COUNT	1668 a	533 c	737 g	889 t
ORIGIN				

Query Match	100.0%	Score 849;	DB 1;	Length 3827;
Best Local Similarity	100.0%	Pred. No. 9,3e-13;		
Matches 849;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	GACATATAAATGGAAAAATCAAAAAGTGACATGATTAAAGCGCATGCCGACAAGC	60
Dd	331	GACGATTAATAATGGAAAAATCAAAAATGGTGCATGTATTAAAGCGCATGCCGACAAGC	390
QY	61	GGTACAGTAAAGATAGAGGGTTATAGTAAAAACAGTACCATTAACTGTTAAAGGTGAACAG	120
Dd	391	GGTACAGTAAAGATAGAGGGTTATAGTAAAAACAGTACCATTAACTGTTAAAGGTGAACAG	450
QY	121	GTGGGTCAAGAGTTATTATCACCCAGACGGTGCACAATTACATTCAATGATTAAGTAGAA	180
Dd	451	GTGGGTCAAGAGTTATTATCACCCAGACGGTGCACAATTACATTCAATGATTAAGTAGAA	510
QY	181	AAATTAAGTGTGTTTTGGGGATTTCGAGAATTTGAAGTACAGGAACAATAATTTAACCAA	240
Dd	511	AAATTAAGTGTGTTTTGGGGATTTCGAGAATTTGAAGTACAGGAACAATAATTTAACCAA	570
QY	241	ACAATACTACTGATGACAAAGTAGTCACGATTAACATCTGGGAATTAATTAACGAATGTT	300
Dd	571	ACAATACTACTGATGACAAAGTAGTCACGATTAACATCTGGGAATTAATTAACGAATGTT	630
QY	301	ACGGTTCATTAAGTGAAGCGGGAACAAGTAGTGTGTTCTATTATTAACCGGAGATATG	360
Dd	631	ACGGTTCATTAAGTGAAGCGGGAACAAGTAGTGTGTTCTATTATTAACCGGAGATATG	690
QY	361	CTACACGAAGATACGACACATGTCAGATGTTTTTAATATATTAACATGAAAAAGTTAT	420
Dd	691	CTACACGAAGATACGACACATGTCAGATGTTTTTAATATATTAACATGAAAAAGTTAT	750

OY	421	GTATCGAAAGTATTACTATATAAAGATCAGATTCCAGGTGGACAGCATGTACATTTAAGC	480
Db	751	GTATCGAAAGTATTACTATATAAAGATCAGATTCCAGGTGGACAGCATGTACATTTAAGC	810
OY	481	ACATTAACAATTAAVTGTGACAGGTACACATATGCAATTTATTATGTGACAAAGTGCATTT	540
Db	811	ACATTAACAATTAAVTGTGACAGGTACACATATGCAATTTATTATGTGACAAAGTGCATTT	870
OY	541	ACGTATTTTGAAAAGCCTTCCACAGTCTCAAAATAAGCTGCATATATACGAAGAACCA	600
Db	871	ACGTATTTTGAAAAGCCTTCCACAGTCTCAAAATAAGCTGCATATATACGAAGAACCA	930
OY	601	ATTGATGTAAACATTTCCACAAAGCTATGGTCAATATATAGTTTTCATTAACACAA	660
Db	931	ATTGATGTAAACATTTCCACAAAGCTATGGTCAATATATAGTTTTCATTAACACAA	990
OY	661	ACCAAAATTTACGAATGAAACAGCAAAAAAGATTTGTTATATATTCACAACTGGTGTCAA	720
Db	991	ACCAAAATTTACGAATGAAACAGCAAAAAAGATTTGTTATATATTCACAACTGGTGTCAA	1050
OY	721	GAGCATGGTAAAGAAAGTGAACGGGAATTCATTATATCATCTGTGCACAAATATTAT	780
Db	1051	GAGCATGGTAAAGAAAGTGAACGGGAATTCATTATATCATCTGTGCACAAATATTAT	1110
OY	781	GCTATGCCGCTATTGAAGTACTGTATAAAGGTGAATTTAAAGTTTTAAACAGATTAA	840
Db	1111	GCTATGCCGCTATTGAAGTACTGTATAAAGGTGAATTTAAAGTTTTAAACAGATTAA	1170
OY	841	GATACCAAG 849	
Db	1171	GATACCAAG 1179	

RESULT	2				PAT	29-SEP-1999
AR067706						
LOCUS	AR067706	3827 bp	DNA			
DEFINITION	Sequence 1 from patent US 5851794.					
ACCESSION	AR067706					
VERSION	AR067706.1	GI:5998928				
KEYWORDS	.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	Unclassified.					
AUTHORS	1 (bases 1 to 3827)					
	Gusé,B., Hook,M., Jonsson,H., Lindberg,M., Patti,J., Sigas,C. and Switalski,L.					
TITLE	Collagen binding protein as well as its preparation					
JOURNAL	Patent: us 5851794-A 1 22-DCC-1998:					
FEATURES	Location/Qualifiers					
SOURCE	1..3827					
	/organism="unknown"					
BASE COUNT	1668 a 533 c 737 g 889 t					
ORIGIN						

Query Match	100.0%	Score 849	DB 6;	Length 3827;
Best Local Similarity	100.0%	Pred NO	9.3e-137;	
Matches 849;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY	1	GACGATAAAAATGCAAAAATGCAAAATGGTACTGATTTAAAGGCGATGCGCCGACAAC	60
Db	331	GACGATAAAAATGCAAAAATGCAAAATGGTACTGATTTAAAGGCGATGCGCCGACAAC	390
QY	61	GGTACAGTAAAGATAGAGGGTTATAGTAAAAAGTACCATTACTGTAAAGGTGAACG	120
Db	391	GGTACAGTAAAGATAGAGGGTTATAGTAAAAAGTACCATTACTGTAAAGGTGAACG	450
QY	121	GTGGGTCAAGCAGTATTATACACACAGCGTGCAACATTCATTCATTAAGTAGAA	180
Db	451	GTGGGTCAAGCAGTATTATACACACAGCGTGCAACATTCATTCATTAAGTAGAA	510
QY	181	AAATTAACTGATGTTGGGATTTGCAGATTGAACTACAGAGAAATTTAAACGCA	240

Db 511 AAATTAAGTATGTTTCGGAGTTTGCAGATTGGAATACAGGAAGAATTTAAGCAA 570
QY 241 ACAAAATCTTCAATGATGCAAGATAGTACATTAACATCTGGGATAATTAACGAATGTT 300
Db 571 ACAAAATCTTCAATGATGCAAGATAGTACATTAACATCTGGGATAATTAACGAATGTT 630
QY 301 AGGTTCAATTAAGTAAAGGAGGAGACAGATAGTGTCTTCAATTAATAAACGGAGATATG 360
Db 631 AGGTTCAATTAAGTAAAGGAGGAGACAGATAGTGTCTTCAATTAATAAACGGAGATATG 690
QY 361 CTACCAAGATAGTACAGATAGTACATGTTTAAATTAATTAACATGAAGAAAGTTAT 420
Db 691 CTACCAAGATAGTACAGATAGTACATGTTTAAATTAATTAACATGAAGAAAGTTAT 750
QY 421 GTATCGAAGATATTAATTAAGATCAGATTCAGAGTGCAGACAGTTAGATTAAAGC 480
Db 751 GTATCGAAGATATTAATTAAGATCAGATTCAGAGTGCAGACAGTTAGATTAAAGC 810
QY 481 ACATTAACATTAATGATGACAGATGACATGATTAATTAATTAATTAATTAATTAAT 540
Db 811 ACATTAACATTAATGATGACAGATGACATGATTAATTAATTAATTAATTAATTAAT 870
QY 541 ACTGATTTGAAAAAGCCTTCCAGTTCTTAAATTAATTAATTAATTAATTAATTAAT 600
Db 871 ACTGATTTGAAAAAGCCTTCCAGTTCTTAAATTAATTAATTAATTAATTAATTAAT 930
QY 601 ATTGATGTAACATTCACAGAGCTATGGTCAATTAATTAATTAATTAATTAATTAAT 660
Db 931 ATTGATGTAACATTCACAGAGCTATGGTCAATTAATTAATTAATTAATTAATTAAT 990
QY 661 ACCAAATTCGATGACAGAGCAAGAAAGTTGTTAATTAATTAATTAATTAATTAAT 720
Db 991 ACCAAATTCGATGACAGAGCAAGAAAGTTGTTAATTAATTAATTAATTAATTAAT 1050
QY 721 GACCATGTAAGAGAGAGTGAACGGGAATTCATTAATTAATTAATTAATTAATTAAT 780
Db 1051 GACCATGTAAGAGAGAGTGAACGGGAATTCATTAATTAATTAATTAATTAATTAAT 1110
QY 781 GCTAATGCGGATTTGAAGTACTGTTAAAGTGAATTAATTAATTAATTAATTAAT 840
Db 1111 GCTAATGCGGATTTGAAGTACTGTTAAAGTGAATTAATTAATTAATTAATTAAT 1170
QY 841 GATACCAAG 849
Db 1171 GATACCAAG 1179

RESULT 3
ALB434 3468 bp DNA PAT 18-APR-1994
LOCUS
DEFINITION Hybrid DNA molecule comprising a nucleotide sequence from S. aureus
coding for a protein or polypeptide having collagen binding
activity.
ACCESSION ALB434
VERSION ALB434.1 GI:513301
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 3468)
AUTHORS
TITLE A COLLAGEN BINDING PROTEIN AS WELL AS ITS PREPARATION
JOURNAL Patent: WO 9207002-A 1 30-APR-1992;
FEATURES Location/Qualifiers
source 1. 3468
/organism="synthetic construct"
/db_xref="taxon:32630"
1. 3465
/note="Protein sequence is in conflict with the conceptual
translation"
/codon_start=1
/transl_table=11
/product="polypeptide having collagen binding activity"

/protein_id="CAU01395.1"
/db_xref="GI:4529908"
/translation="ARDISTVNTDLTSPKIEDGKTYVKTEDDKNGKIONGMI
KVAMPSTGVKIEGSKVPLTVKGEQVQVQATTPGCAITLNDKIEKLSVSGEAF
EVOGRNLQDTNLDKDKVATISGKSTANIGIKVREVPVILINKSGICVQEDETH
VWFLINNEKSYVSKDITIKDIOGGQOLDLNIINTVGHSHSNVSCQSLATFEK
APPGSKITVDNFKNTIDVTIIPQGVSGSNVYKIKINTEQKVEVNSQAWYOHG
KEEVKSPFNHYHINANAGIEGYKSGELKVKOKDKAPIANVFKLSKKDSVY
KDNQKEIETITPANGIANIKALPSGDYILKEIAPRPVFDKDEKPTPMKIDNOGY
GTTIENAKAIEKTDVSAQWEGDQVAPPTIYFKLYKODDNQNTVPVKAELKLED
GTTKIVSNMLPEKDKNGKAIKYLKVENAOGEDTPEGYTKKENGVLNTEKPIETT
SISGEKVMDDKDNODKREPKSVNLANGEKVIDVTSENMKEFEFDLPYDEK
KIEVYTEDHVDKDTDINGTITNKPYPGSETSAVTYKMMDDNNODKREPKSVL
YODGKATGKTATLINESNMNTHTYTGDEKAKQOYKYIYBELTKYGTTHYDNDNG
NLIVTKYTPETTSISGEKVMDDKDNODKREPKSVNLADSEKVIDVTSENMK
YEKDLKPYDEKREKIEVYTEDHVDKDTDINGTITNKPYPGSETSAVTYKMMDDNN
ODGKREPEIKVELYODGKATGKTATLINESNMNTHTYTGDEKAKQOYKYIYBELTKY
KGTTHYDNDNGNLIVTKYTPETTSISGEKVMDDKDNODKREPKSVNLANGEK
VKTIDVTSSTNNKYEKDLKPYDEKREKIEVYTEDHVDKDTDINGTITNKPYPG
SATVTYKMMDDNNODKREPEIKVELYODGKATGKTATLINESNMNTHTYTGDEKAK
QOYKYIYBELTKYGTTHYDNDNGNLIVTKYTPETTSISGEKVMDDKDNODKREPKSV
HSNKVPTPEPKRPSKVDKDDPKDNKTKPENLKLKPKYKIMTITWTFIGILGL
YLIRKRFN"

BASE COUNT 1532 a 506 c 689 g 741 t
ORIGIN
Query Match 93.3%; Score 792.2; DB 6; Length 3468;
Best local Similarity 97.8%; Pred. No. 5.3e-127;
Matches 836; Conservative 0; Mismatches 13; Indels 6; Gaps 3;
QY 1 GACGATTAATTAAGTGAAGAAATGATGACATGATTTAAAGTGATGCGGACAGC 60
Db 94 GACGATTAATTAAGTGAAGAAATGATGACATGATTTAAAGTGATGCGGACAGC 153
QY 61 GGTACAGTAAGATAGAGGTTATAGTAAAGACGTACATTAATCTTTAAAGTGAAG 120
Db 154 GGTACAGTAAGATAGAGGTTATAGTAAAGACGTACATTAATCTTTAAAGTGAAG 213
QY 121 GTGGGTCAGACAGTTATTACACAGAGCGTGCAACATTAATCAATTCATTAAGTGA 180
Db 214 GTGGGTCAGACAGTTATTACACAGAGCGTGCAACATTAATCAATTCATTAAGTGA 273
QY 181 AAATTAAGTATGTTGGGATTTGCGAATTTGAAGTACAGAGAAATTTAAGC 240
Db 274 AAATTAAGTATGTTGGGATTTGCGAATTTGAAGTACAGAGAAATTTAAGC 333
QY 241 ACAATTAATCTTCAATGATGCAAGATAGTACATTAATTAATTAATTAATTAAT 300
Db 334 ACAATTAATCTTCAATGATGCAAGATAGTACATTAATTAATTAATTAATTAAT 393
QY 301 A--CGGTTCAATTAAGTGAAGCGGAGACAGTGTGTTTCTATTATTAATA--CGGAG 355
Db 394 ATCGGTTGATTAAGTGAAGCGGAGACAGTGTGTTTCTATTATTAATAAGCGGAG 453
QY 356 ATATGCTACC-AGAAAGTACGACACATGTACAGATGTTTAAATTAATTAAGTGA 414
Db 454 ATATGCTACCAGAAAGTACGACACATGTACAGATGTTTAAATTAATTAAGTGA 513
QY 415 AGTTATGATGCAAGAGTATTATTAAGATGATTAAGTGAAGTGAAGCGGATTTGAT 474
Db 514 AGTTATGATGCAAGAGTATTATTAAGATGATTAAGTGAAGTGAAGCGGATTTGAT 573
QY 475 TTAAAGCATTAATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 534
Db 574 TTAAAGCATTAATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 633
QY 535 GCAATTAATGATTTGAAAAAGCCTTCCAGGTTCTAATAATTAATCTGTTGATTAAG 594
Db 634 GCAATTAATGATTTGAAAAAGCCTTCCAGGTTCTAATAATTAATCTGTTGATTAAG 693
QY 595 AACACATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 654

```

Db 694 AACACATTGATGTAACAAATCCACAAGGCTATGGCTCATATAATAGTTTTCATTAATTAAC 753
QY 655 TACAAAACCAAAATTAAGATGAAGACAGAAAAAGAGTTGGTAATAATTAACAAAGCTGG 714
Db 754 TACAAAACCAAAATTAAGATGAAGACAGAAAAAGAGTTGGTAATAATTAACAAAGCTGG 813
QY 715 TATCAAGAGCATGTTAAGGAAGAGTGAACGGGAATCATTTAATCATCTGTCACAAAT 774
Db 814 TATCAAGAGCATGTTAAGGAAGAGTGAACGGGAATCATTTAATCATCTGTCACAAAT 873
QY 775 ATTAATGCTAATCCGGTATTTGAAGTACTGTAAAGGTGAATTAAGTTTAAAGCAAG 834
Db 874 ATTAATGCTAATCCGGTATTTGAAGTACTGTAAAGGTGAATTAAGTTTAAAGCAAG 933
QY 835 GATTAAGATACCAAG 849
Db 934 GATTAAGATACCAAG 948

RESULT 4
A18436 4612 bp DNA PAT 18-APR-1994
LOCUS assembled sequence from the insert in p 16 and cCOLR6A.
DEFINITION A18436
ACCESSION A18436
VERSION A18436.1 GI:513303
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 4612)
AUTHORS
TITLE A COLLAGEN BINDING PROTEIN AS WELL AS ITS PREPARATION
JOURNAL Patent: WO 9207002-A 3 30-APR-1992;
FEATURES
source
1. 4612
/organism="synthetic construct"
/db_xref="taxon:32630"
/codon_start=1
/product="polypeptide with collagen binding activity"
/protein_id="CA01396.1"
/db_xref="GI:513304"
/translation="MKNVNLKEMVFMILNITPLEFNKNEAFARDISSTNVTLTFS
PKRIEDGKTYVMKTFDDKNGKIONGDMKIVAMPISGTVKIEGSKTVPLTVKGEQV
QAVITPDGATITFENDKVEKLSVSGFAEFVGGRLNLTOTLDDKVAITTSNKSSTNV
IGMIVKREPVEVLINSGKICGYOEDTHVRFMLNNKSEYVSKDITTKDIOGGQO
LILSTLINIVTGHSHNYSGOSALIDPEKAFPGSKTYTNDTNTDVTIPDGYGSNS
ESINRKTITNNOQKRFVNSQAMTOEHKEKEVNGKSEPHYTHNINAAGISCTYKGE
LAVLQODKTPKAPLANVFKLISKDSVVKDOKETELITDANGIANIKALPSGDYIL
KEIEARPYTFDEKDEYPTMKDTONOGVFTIENKALEKRDVSAQVWEGTOYK
PIYFELVYKODNONMTTPVDKAEIKKLEGGTKVWSNLPENDKNGKAIKYVKEVNA
OGEEDTPEGYTKKENGIVYTNTEKPIETTSISEGKWDKDKDNDKRPKRPVSNILAN
GKSVKTLDTSETNMKIEFKDLPKYDEGKIEYTVEDHVKYTTDINGTTITNKYTP
GFSATVPMKNDNNNOGDKRPEIKVELYQOGKATGKATPLINESNNHTHTTGDEK
AKGQOVKTYVEELTKVKGITTVNDNDMGNLVTNKTYPETTSISEGKWDKDKDNDG
KRPKRPVSNILADGEKVTLDVTSNTMYEKRPDKYDEGKIEYTVEDHVKYTT
DINGTTITNKYTPGFSATVPMKNDNNNOGDKRPEIKVELYQOGKATGKATPLINES
NNMTHTWGLDEKAKGOOVKYVEELTKVKGITTVNDNDMGNLVTNKTYPETTSIS
GKVMDDKDNODGKRPKRPVSNILANGKELVTLDTSETNMKIEFKDLPKYDEGKIE
YTVEDHVKYTTDINGTTITNKYTPGFSATVPMKNDNNNOGDKRPEIKVELYQO
GKATGKATPLINESNNMTHTWGLDEKAKGOOVKYVDELTKVNGYTHVNDNDMGNLI
VNTKPKRPKRPVPEKRPDKTPPTPKPHSKNPKVPTPDKPSKVDKDPDKNTKTP
ENPLKELPTGKMKIITSWITWFIIGILIGILYILIRKRFNS"

BASE COUNT 1946 a 676 c 868 g 1122 t
ORIGIN

```

```

Query Match 93.3%; Score 792.2; DB 6; Length 4612;
Best Local Similarity 97.8%; Pred. No. 5,1e-127;
Matches 836; Conservative 0; Mismatches 13; Indels 6; Gaps 3;
QY 1 GACGATRAAAAATGGAATAATGATGACATGATTAAGTGGCAGCCGACAAAGC 60

```

```

Db 1111 GACGATRAAAAATGGAATAATGATGACATGATTAAGTGGCAGCCGACAAAGC 1170
QY 61 GGTACAGTAAGATAGAGGTTTATAGTAACAGTACCAATTAAGTGTAAAGGTGAACAG 120
Db 1171 GGTACAGTAAGATAGAGGTTTATAGTAACAGTACCAATTAAGTGTAAAGGTGAACAG 1230
QY 121 GTGGGTCAAGCAGTATTAACACAGAGCGTGCAACAATTAATCAATGATTAAGTAAGAA 180
Db 1231 GTGGGTCAAGCAGTATTAACACAGAGCGTGCAACAATTAATCAATGATTAAGTAAGAA 1290
QY 181 AATTAAGTATGATGTTTCGGGATTTGCAAAATTTGAATGCAAGGAAGAAATTAAGCAAA 240
Db 1291 AATTAAGTATGATGTTTCGGGATTTGCAAAATTTGAATGCAAGGAAGAAATTAAGCAAA 1350
QY 241 ACAATTAATCTTACAGTACGACAAAGTAGTACGATTAACATCTGGGAATTAATCAAGTAATGTT 300
Db 1351 ACAATTAATCTTACAGTACGACAAAGTAGTACGATTAACATCTGGGAATTAATCAAGTAATGTT 1410
QY 301 A--CGGTTCAATAAAGTGAAGCGGGAACAAGTAGTGTCTTATTAATAA---CGGGAAG 355
Db 1411 ATCGGTTGATTAAGTGAAGCGGGAACAAGTAGTGTCTTATTAATAAAGCGGGAAG 1470
QY 356 ATATGCTAGC-AGAAATACGACATGTAATGATGTTTAAATATTAACAATGAAGAA 414
Db 1471 ATATGCTACCAAGAAAGATACGACATGTAATGATGTTTAAATATTAACAATGAAGAA 1530
QY 415 AGTTATGATGCAAGAAATATTAATTAAGGAATCAATTAAGGAATCAAGTACAGTATGAT 474
Db 1531 AGTTATGATGCAAGAAATATTAATTAAGGAATCAATTAAGGAATCAAGTACAGTATGAT 1590
QY 475 TTAAAGCATTAAACATTAATGATGACAGGTACACATAGCAATTAATTAAGTGACAAAGT 534
Db 1591 TTAAAGCATTAAACATTAATGATGACAGGTACACATAGCAATTAATTAAGTGACAAAGT 1650
QY 535 GCAATTAAGTATTTGAAAAAGCGTTTCCAGGTTCTAAATTAATCTGTGATTAATAGCAAG 594
Db 1651 GCAATTAAGTATTTGAAAAAGCGTTTCCAGGTTCTAAATTAATCTGTGATTAATAGCAAG 1710
QY 595 AACACATTGATGTAACAAATTCACAAAGGCTATGGGTCATTAATTAAGTTTCAATTAAC 654
Db 1711 AACACATTGATGTAACAAATTCACAAAGGCTATGGGTCATTAATTAAGTTTCAATTAAC 1770
QY 655 TACAAAACCAAAATTAAGATGAAGACAGAAAAAGAGTTGGTAATAATTAACAAAGCTGG 714
Db 1771 TACAAAACCAAAATTAAGATGAAGACAGAAAAAGAGTTGGTAATAATTAACAAAGCTGG 1830
QY 715 TATCAAGAGCATGTTAAGGAAGAGTGAACGGGAATCATTTAATCATCTGTCACAAAT 774
Db 1831 TATCAAGAGCATGTTAAGGAAGAGTGAACGGGAATCATTTAATCATCTGTCACAAAT 1890
QY 775 ATTAATGCTAATCCGGTATTTGAAGTACTGTAAAGGTGAATTAAGTTTAAAGCAAG 834
Db 1891 ATTAATGCTAATCCGGTATTTGAAGTACTGTAAAGGTGAATTAAGTTTAAAGCAAG 1950
QY 835 GATTAAGATACCAAG 849
Db 1951 GATTAAGATACCAAG 1965

RESULT 5
AR067712
LOCUS AR067712 4612 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 8 from patent US 5851794.
ACCESSION AR067712
VERSION AR067712.1 GI:5998934
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 4612)
Guss, B., Hook, M., Jonsson, H., Lindberg, M., Patil, J., Signas, C. and
SwitalSKI, L.

```


[illegible]

RESULT	7
AC084397/c	
LOCUS	AC084397 171798 bp DNA HTG 17-JUL-200
DEFINITION	Trypanosoma brucei chromosome unknown clone RCI193-II18, ***
SEQUENCING	SEQUENCING IN PROGRESS ***, 1 ordered pieces.
ACCESSION	AC084397
VERSION	AC084397
KEYWORDS	AC084397.4 GI:11612622
SOURCE	HTG: HTGS_PHASE2.
ORGANISM	Trypanosoma brucei. Trypanosoma brucei

REFERENCE	AUTHORS	TITLE	JOURNAL
1 (bases 1 to 17198)	El-Sayed, N.M., Ghedin, E., Song, J., Larkin, C., Wanless, D., Jones, K., Peterson, J., Hou, L., Zhao, H., Mason, T., Miltischer, J., Pai, G., Van Aken, S., Uterback, T., Khalak, H.G., Gerard, C., Leech, V., Ull, E., Melville, S., White, O., Adams, M.D., Donaldson, J.E. and Fraser, C.M.	Trypanosoma brucei GUTat10.1 RPL193-1118 BAC genomic sequence	Unpublished
2 (bases 1 to 17198)	El-Sayed, N.M., Khalak, H. and Adams, M.D.	Direct Submission	Submitted (30-OCT-2000)

```

COMMENT
on Dec 11, 2000 this sequence version replaced g1:11230847.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 171798: contig of 171798 bp in length.

```

```

FEATURES
Source
1. 171798
location/Qualifiers
/organism="Trypanosoma brucei"
/isolate="GUTat10.1"
/db_xref="taxon:5631"
/chromosome="unknown"
/clone="RPC193-1118"
BASE COUNT
39792 a 37575 c 33263 g 61168 t
ORIGIN

```

Query Match	Score	DB 2	Length
7.7%	65.6	171798	

Best Local Similarity 46.8%; Pred. NO. 0.016;
Matches 206; Conservative 0; Mismatches 234; Indels 0; Gaps 0

[illegible]

OY	457	GGTGACAGCGCTTGAAGTTTAAGCAATTTAAACTTTAATGGACAGGTACAACATTCAGT	516
		" " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " "	
Dd	158664	AATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	158605
OY	517	TATTATAGTGCAACAAGTGCATTTACTGATTTTGGAAAAACCCTTCCAGGTTCTAAAATA	576
		" " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " "	
Dd	158604	AATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	158545
OY	577	ACTGTTGATTAATACGAAGAACAATTTGATGTAACAATTTCCACAAGGCTTAGGGTCATAT	636
		" " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " "	
Dd	158544	AATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	158485
OY	637	AATAGTTTTCAATTAACTACACCAAACCAAAATTTCCGAATGCACGCAAAAGAGTTTGT	696
		" " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " "	
Dd	158484	AATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	158425
OY	697	AATAATTCACAAGCTTGSTATCAAGACATGCTAAGAGNAGTGAACGGGAATCATTT	756
		" " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " "	
Dd	158424	AATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	158365
OY	757	AATCATTCTGTGCCAATATTAAATGCTAAATGCCGGTATTGAAGTACTGTAAAAGGTGAA	816
		" " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " "	
Dd	158364	AATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	158305
OY	817	TTAAACGTTTAAAAACAGA	836
		" " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " "	
Dd	158304	GTACTAATATTATTAAGTATGA	158285

RESULT		8					PAT		28-DEC-1997
LOCUS	166494/C								
DEFINITION	Sequence 14 from patent US 5670367.								
ACCESSION	166494		7218 bp	DNA					
VERSION	166494.1								
KEYWORDS	GI:2724471								
SOURCE	.								
ORGANISM	Unknown.								
	Unknown.								
REFERENCE	unclassified.								
AUTHORS	1 (bases 1 to 7218)								
TITLE	Dorner, F., Scheifflinger, F. and Falkner, F. Gunter.								
JOURNAL	Recombinant fowlpox virus Patent: US 5670367-A 14 23-SEP-1997;								
FEATURES	Location/Qualifiers								
source	1..7218								
BASE COUNT	/organism="unknown"								
ORIGIN	1944 a 1491 c 1486 g 1929 t 368 others								

Query Match	7.68;	Score 64.2;	DB 6;	Length 7218;
Best Local Similarity	6.58;	Pred. NO. 0.042;		
Matches 27;	Conservative 225;	Mismatches 163;	Indels 0;	Gaps 0

[illegible]

QY 401 TTAACAATGAAAGTTATGATCGAAGATATATCTAATAAGATCAGATCAAGTG 460
DB 1174 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1115
QY 461 GACAGCAGTATGATTTAAGCAGATTAACATTAATGTCAGAGTACACATAGCAA 515
DB 1114 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1060
RESULT 9
AC007926 155204 bp DNA HTG 17-JUL-2001
LOCUS Trypanosoma brucei chromosome II clone RCI93-3H15, *** SEQUENCING
DEFINITION IN PROGRESS ***, 1 ordered pieces.
AC007926 GI:14787210
VERSION HTG: HTGS_PHASE2.
KEYWORDS Trypanosoma brucei.
SOURCE Trypanosoma brucei.
ORGANISM Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 155204)
AUTHORS El-Sayed,N.M., Ghedira,E., Song,J., Larkin,C., Manless,D., Jones,K.,
Peterson,J., Hou,L., Zhao,H., Mason,T., Mlitscher,J., Pal,G., Van
Aken,S., Uterback,T., Khalak,H.G., Gerard,C., Leech,V., Ullu,E.,
Meylille,S., White,O., Adams,M.D., Donelson,J.E. and Fraser,C.M.
Trypanosoma brucei GUTat10.1 RCI93-3H15 BAC genomic sequence
Unpublished
2 (bases 1 to 155204)
AUTHORS El-Sayed,N.M., Khalak,H. and Adams,M.D.
DIRECT SUBMISSION
Submitted (28-JUN-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Jul 17, 2001 this sequence version replaced gi:12746529.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 155204: contig of 155204 bp in length.
FEATURES
source
1.155204
/organism="Trypanosoma brucei"
/isolate="GUTat10.1"
/db_xref="taxon:5691"
/chromosome="II"
/clone="RCI93-3H15"
BASE COUNT 48747 a 31603 c 32329 g 42525 t
ORIGIN
Query Match 7.4%; Score 63.2; DB 2; Length 155204;
Best Local Similarity 46.2%; Pred. No. 0.041;
Matches 209; Conservative 0; Mismatches 243; Indels 0; Gaps 0;
QY 397 AATATTAACATGAAAAAGTTATGATCGAAGATATCTAATAAGATCAGATCA 456
DB 53570 AATTAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAAT 53629
QY 457 GGTGACAGCAGTATGATTTAAGCAGATTAACATTAATGTCAGAGTACATAGCAAT 516
DB 53630 AATTAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAAT 53689
QY 517 TATTAATAGTCGACAAAGTCAATTAATGATTTGAAAAAGCCCTTCCAGGTTCTAAATA 576
DB 53690 AATTAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAAT 53749
QY 577 ACTGTGATTAATAAGAAACAAATGATGTAACAATCCACAAGGCTATGGGTATAT 636

DB 53750 AATTAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAAT 53809
QY 637 AATAGTTTTCATTAACACAAAACCAAAATTAAGAAAGCAAGCAAAAAGACTTGT 696
DB 53810 AATTAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAAT 53869
QY 697 AATTAATCAGAGTTGCTATCAAGACATGCTAAGCAAGAAAGCAAGCAAGTAT 756
DB 53870 AATTAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAAT 53929
QY 757 AATCATCTGTCGACAAATTAATGTAATGCCGTTATGAGCTAGTGTAAAGTCAA 816
DB 53930 AATTAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAAT 53989
QY 817 TTAAGAGTTTAAACAGATTAAGATACCAA 848
DB 53990 TACTAATATTAATAATAAGACAGTAATAATAA 54021
RESULT 10
AC083827 208546 bp DNA HTG 02-OCT-2000
LOCUS Homo sapiens clone Rp11-345C21, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC083827
AC083827 GI:10445290
VERSION HTG: HTGS_PHASE0.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 208546)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 208546)
AUTHORS Waterston,R.H.
DIRECT SUBMISSION
Submitted (02-OCT-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
Center project name: H_NH0345C21.
* NOTE: This record contains 253 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 1642: contig of 1642 bp in length
1643: gap of unknown length.
1653 2130: contig of 478 bp in length
2131 2140: gap of unknown length
2141 3371: contig of 1231 bp in length
3372 3381: gap of unknown length
3382 4173: contig of 792 bp in length
4174 4184: gap of unknown length
4185 5657: contig of 1474 bp in length
5658 5667: gap of unknown length
5668 6470: contig of 803 bp in length
6471 6480: gap of unknown length
6481 7499: contig of 1019 bp in length
7500 7510: gap of unknown length
7511 8801: contig of 1291 bp in length
8802 8810: gap of unknown length
8811 9274: contig of 464 bp in length
9275 9285: gap of unknown length
9286 10198: contig of 914 bp in length
10199 10208: gap of unknown length

*	10209	11002: contig of 794 bp in length
*	11003	11012: gap of unknown length
*	11013	12073: contig of 1061 bp in length
*	12074	12083: gap of unknown length
*	12084	12197: contig of 114 bp in length
*	12198	12307: gap of unknown length
*	12208	13363: contig of 1156 bp in length
*	13364	13373: gap of unknown length
*	13374	14175: contig of 802 bp in length
*	14176	14185: gap of unknown length
*	14186	15448: contig of 1263 bp in length
*	15449	15458: gap of unknown length
*	15459	16424: contig of 966 bp in length
*	16425	16434: gap of unknown length
*	16435	17475: contig of 1041 bp in length
*	17476	17485: gap of unknown length
*	17486	18253: contig of 768 bp in length
*	18254	18263: gap of unknown length
*	18264	19283: contig of 1020 bp in length
*	19284	19293: gap of unknown length
*	19294	20134: contig of 841 bp in length
*	20135	20144: gap of unknown length
*	20145	21630: contig of 1486 bp in length
*	21631	21640: gap of unknown length
*	21641	22167: contig of 527 bp in length
*	22168	22177: gap of unknown length
*	22178	23135: contig of 958 bp in length
*	23136	23145: gap of unknown length
*	23146	24505: contig of 1360 bp in length
*	24506	24515: gap of unknown length
*	24516	26306: contig of 1791 bp in length
*	26307	26316: gap of unknown length
*	26317	27128: contig of 812 bp in length
*	27129	27138: gap of unknown length
*	27139	28060: contig of 922 bp in length
*	28061	28070: gap of unknown length
*	28071	29460: contig of 1390 bp in length
*	29461	29470: gap of unknown length
*	29471	30856: contig of 1386 bp in length
*	30857	30866: gap of unknown length
*	30867	31404: contig of 538 bp in length
*	31405	31414: gap of unknown length
*	31415	32425: contig of 1011 bp in length
*	32426	32435: gap of unknown length
*	32436	33719: contig of 1284 bp in length
*	33720	33729: gap of unknown length
*	33730	34833: contig of 1094 bp in length
*	34834	34833: gap of unknown length
*	34834	35826: contig of 993 bp in length
*	35827	35836: gap of unknown length
*	35837	36653: contig of 817 bp in length
*	36654	36663: gap of unknown length
*	37334	37333: contig of 670 bp in length
*	37344	37343: gap of unknown length
*	37344	37531: contig of 188 bp in length
*	37532	37541: gap of unknown length
*	37542	38960: contig of 1419 bp in length
*	38961	38970: gap of unknown length
*	38971	39778: contig of 808 bp in length
*	39779	39788: gap of unknown length
*	39789	41572: contig of 1784 bp in length
*	41573	41582: gap of unknown length
*	41583	42900: contig of 1318 bp in length
*	42901	42910: gap of unknown length
*	42911	43747: contig of 837 bp in length
*	43748	43757: gap of unknown length
*	43758	44981: contig of 1224 bp in length
*	44982	44991: gap of unknown length
*	44992	45495: contig of 504 bp in length
*	45496	45505: gap of unknown length
*	45506	46565: contig of 1060 bp in length
*	46566	46575: gap of unknown length
*	46576	47763: contig of 1188 bp in length

*	47764	47773: gap of unknown length
*	47774	49126: contig of 1353 bp in length
*	49127	49136: gap of unknown length
*	49137	49976: contig of 840 bp in length
*	49977	49986: gap of unknown length
*	49987	51259: contig of 1273 bp in length
*	51260	51269: gap of unknown length
*	51270	51810: contig of 541 bp in length
*	51811	51820: gap of unknown length
*	51812	53886: contig of 2066 bp in length
*	53887	53896: gap of unknown length
*	53897	54709: contig of 813 bp in length
*	54710	54719: gap of unknown length
*	54720	55769: contig of 1050 bp in length
*	55770	55779: gap of unknown length
*	55780	56320: contig of 541 bp in length
*	56321	56330: gap of unknown length
*	56331	57713: contig of 1383 bp in length
*	57714	57723: gap of unknown length
*	57724	58504: contig of 781 bp in length
*	58505	58514: gap of unknown length
*	58515	59576: contig of 1062 bp in length
*	59577	59586: gap of unknown length
*	59587	60391: contig of 805 bp in length
*	60392	60401: gap of unknown length
*	60402	61616: contig of 1215 bp in length
*	61617	61626: gap of unknown length
*	61627	63058: contig of 1432 bp in length
*	63059	63068: gap of unknown length
*	63069	64355: contig of 1287 bp in length
*	64356	64365: gap of unknown length
*	64366	64696: contig of 334 bp in length
*	64700	64709: gap of unknown length
*	64710	65543: contig of 834 bp in length
*	65544	65553: gap of unknown length
*	65554	66485: contig of 932 bp in length
*	66486	66495: gap of unknown length
*	66496	66743: contig of 248 bp in length
*	66753	66753: gap of unknown length
*	66754	67732: contig of 979 bp in length
*	67733	67742: gap of unknown length
*	67743	68475: contig of 733 bp in length
*	68476	68485: gap of unknown length
*	68486	69665: contig of 1180 bp in length
*	69666	69675: gap of unknown length
*	69676	70517: contig of 842 bp in length
*	70518	70527: gap of unknown length
*	70528	71693: contig of 1166 bp in length
*	71694	71703: gap of unknown length
*	73082	73081: contig of 1378 bp in length
*	73092	73091: gap of unknown length
*	73917	73916: contig of 825 bp in length
*	73917	73926: gap of unknown length
*	73927	74769: contig of 843 bp in length
*	74770	74779: gap of unknown length
*	74780	75081: contig of 302 bp in length
*	75082	75091: gap of unknown length
*	75092	75731: contig of 640 bp in length
*	75732	75741: gap of unknown length
*	75742	76481: contig of 740 bp in length
*	76482	76491: gap of unknown length
*	76492	77493: contig of 1002 bp in length
*	77494	77503: gap of unknown length
*	77504	78239: contig of 736 bp in length
*	78240	78249: gap of unknown length
*	78250	79249: contig of 1000 bp in length
*	79250	79259: gap of unknown length
*	79260	79751: contig of 492 bp in length
*	79752	79761: gap of unknown length
*	79762	80611: contig of 850 bp in length
*	80612	80621: gap of unknown length
*	80622	81941: contig of 1320 bp in length
*	81942	81951: gap of unknown length

OY	4	GAAAAAATGGAAAAATTCAAAATGCGTCACTGATTAATAGGCGATGGCCGCAACGGGT	63
Db	63352	GATTAATAATTAATTTAAATTAATCTGAAAATTTAGAAAACAAATGTTCTTCAGGATCAT	63411
OY	64	ACAGTAAAGATAGAGGGTTATAGTAAACACAGTACCATTACTGTTTAAAGGGTGAACGGTGG	123
Db	63412	AATATTATTAATGAAGAAAATTTAAAAAATTTGAAAAAAGAAAAAAGAGATTATG	63471
OY	124	GGTCAAGCAGTTATTACACCGACGGGTCAACAAATTACATTCATGATTAAGTAGAAAA	183
Db	63472	TTTTCCAGTATTACTTGAAGAAATATGCGCAAGAAAATAATTCAAAAGTCAAAAAAAGCAT	63531
OY	184	TTAAAGTATGTTTGGCGATTTGCAGAAATTTGAGTACAGAAAGAAATTTAAACGCAACA	243
Db	63532	ACAAAGCAATGTTATGACACACATATTTAAATGAAAAAATGTTAAATGAATTAATCTTTTCA	63591
OY	244	AATACTTCAGATGACAAAAGTAGACTACGATTAACATCTGGGAATTAATTCACAGATTTACG	303
Db	63592	AATGAATCAAAAATTAATGGATGAAGAAAATTTCCAAATTAGAAAAACAAAATTAATCTTTA	63651
OY	304	GTTCTAAAGTGAAGCGGGAACAAGTAGTGTTCATTTATTAATACGGAGATATGCTA	363
Db	63652	CAAAATTAATAAATGGAGATGATTAACGAAAGATTAATAAATTAACAAAAAATAATATGTTT	63711
OY	364	CCAGAGATGATGACACATGTAAGATGTTTAAATTAATCAAGAAAAAGTATATGTA	423
Db	63712	TCAGTGTGATGTCAAAATATTTTAATGAGATTTCAAAATTAATGAATAAAGAAAAAGGTGAAA	63771
OY	424	TGCAAGATATTACTATTAAGATGATGATTCACAGGTGACAGCAGTGTGATTTAAGACA	483
Db	63772	AAAAAAGGGTACTTTTATGATGAATTAAGAAAGGACAAAAAGTTGATGATCCAAA	63831
OY	484	TTAAACATTATGTGACAGGTACACATAGCAATTTATTAAGTGACAAAGTGCATTTACT	543
Db	63832	AAAAATTAACATTTGAAAAAATGATATATATCAATATATACAAAAAATTAATATATAG	63891
OY	544	GATTTTGGAAAAACCTTTCCAGGTTCTTAAATTAACATGTTGATTAATAGCAAGAACCAATT	603
Db	63892	AATTCGAAAAAAGATGAACAAAGTACGGTACTATTTCTTAGATGAGCTCAAAAAAGACACA	63951
OY	604	GATGTA---ACAAATTCACAAGGCTATGCGTCATTAATAGTTTTCATTAATACATACAAA	660
Db	63952	AATGTAATTAATAGTGAATTAATATCTTTTCAAAATCAACCTGGAATTAATTTACGCAAT	64011
OY	661	ACCAAAATTCAGATGAACACGCAAAAAGATTTGTTAAATTAATTCACAAAGCTTGGATCAA	720
Db	64012	AACATTAAGATGATTAATGAAAAATTAATATGATTTAAAAAATATATATTGGAAAGAAAGA	64071
OY	721	GAGCATGGTAAGGAAGAGTGA	743
Db	64072	AAAGAAAGAAATCAAGAGTAAA	64094

RESULT	14
LOCUS	AC004157
DEFINITION	AC004157 169546 bp DNA HTG 12-AUG-2000
ACCESSION	AC004157
VERSION	AC004157.8
KEYWORDS	GI:9797712
SOURCE	HTG: HTGS-PHASE1.
ORGANISM	malaria parasite <i>P. falciparum</i> .
REFERENCE	Plasmodium falciparum
AUTHORS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
TITLE	1 (bases 1 to 169546)
JOURNAL	Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,
REFERENCE	Kutdi,O.B., Conway,A.B. and Davis,R.W.
AUTHORS	Plasmodium falciparum 3D7 chromosome 12
TITLE	unpublished
JOURNAL	2 (bases 1 to 169546)
REFERENCE	Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
AUTHORS	Direct Submission
TITLE	Submitted (19-FEB-1998) Stanford DNA Sequencing and Technology

Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
On Aug 12, 2000 this sequence version replaced gi:8810447.

On Aug 12/2000 this sequence version replaced q1:8610447
 NOTE: This is a 'working draft' sequence. It currently
 consists of 2 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

FEATURE

*	1	23466: contig of 23466 bp in length
*	23467	23666: gap of unknown length
*	23667	169546: contig of 145880 bp in length
	Location/Qualifiers	
	1..169546	

BASE COUNT	69871	a	15381	c	15705	g	68389	t	200	others
ORIGIN										

Query Match	7.28;	Score 61.4;	DB 2;	Length 169546;
-------------	-------	-------------	-------	----------------

Matches 324; Conservative 0; Mismatches 416; Indels 3; Gaps 1;

OY	4	GATTAATAATGCAAAATACAAAATGCTGACATGATTAAGAATGGCATGGCCGACAGCGT	63
Db	74936	GATTAATAATTTAATTTGAAATATCTTGAAATTAATGAAAACAAGTTCTTACGATGAT	74995
OY	64	ACAGTAAAGATAGAGGGTTATAGTAAAAACAGTACCATTAAGTTTAAAGTGAACAGTG	123
Db	74996	AATATTAATTAATGAGAAAATTTTAAAAAATATGAAAAAAGAGAGTTATG	75055
OY	124	GGTCAGCAGTATATACACCGGTCGAACAATTAAGTTCAATGATTAAGTAGAAAA	183
Db	75056	TTTTCAGATATATCTTGAAAGAAATATGGCAAGAAATATTCAAATGTGAAAAAACAAT	75115
OY	184	TTTAAGTGAATTTCCGGATTTCCACAATTTGGAATGACAGGAAGAAATTTACGCAACA	243
Db	75116	ACAAACAAATGTTATGCAACACATATTAATGATGAAAAATGTTAAATGAATTAAGTTTCA	75175
OY	244	AATACTTCAGATGACAAAGTACCTACGATACATCTGGGAATTAATCAACGAATGTTACG	303
Db	75176	AATGATACAAAATTAATGGATGACAAAAATATCCAAATTTGAAAAACAATAATATCTTTA	75235
OY	304	GTTCAATTAAGTGAAGCGGAGCAAGTAGTGTCTTCTATTAATAAAGGGAATATGCTA	363
Db	75236	CAAAATTAATAATGAGATGATTAACGAAAGATTAATAACAAAAAATAATTAATGTTT	75295
OY	364	CCAGAGATACGACACATGTACGATGTTTTTAATATTAACATGAAAAAAGTTATGTA	423
Db	75296	TCAGATGATGTCAAAATTTTATAGAGATTCGAATTAATGAAAAATAGAAAAAGTGAAA	75355
OY	424	TCGAAAGATATTACTATTAAGATACAGATTCAAGGTGGACAGCAGTTAGATTAAACACA	483
Db	75356	AAAAAAGGGTAACTTTATGATGATCAATTAGAAGAGGACAAAAAGTTTGATCCAAAA	75415
OY	484	TTAAACATTAAATGTGACAGGTACATACATCAATTTTATATGAGCAAAAGTCAATTACT	543
Db	75416	AAAAATTAACATTGAAAAAATATATATATATCTTAATAGCAAAAAATATATATATATG	75475
OY	544	GATTTTGA AAAAGCCTTTCCAGTTCTAAATTAAGTTGATTAATGCAAGACAAT	603
Db	75476	AATGACAAAAAAGATGACAAAGTACGGTACTATCTTGATGACGTCAAAAAAGACAA	75535
OY	604	GATGTA---ACAATTCACACAGGCTATGGGTATTAATAGTTTTCATTAATGACAA	660
Db	75536	AATGATATTAATAGATTAATAATATCTTTCAAAATCAACCTGAAATATATTTACTCAAT	75595

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2002, 21:25:46 ; Search time 1549.43 Seconds
(without alignments)
5888.079 Million cell updates/sec

Title: US-09-813-820-3
Perfect score: 849
Sequence: 1 GACGATATAAATGCAAAAT.....AACAGATTAAGATACCAAG 849

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estda:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_iny:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rpd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	72.4	8.5	1007	13	CNS06X9S
2	67.4	7.9	1101	13	CNS00EVL
3	67	7.9	641	13	AO946120
4	65.2	7.7	1101	13	CNS0039G
5	64.8	7.6	1101	13	CNS0182P
6	64.6	7.6	1059	13	CNS0022B
7	64	7.5	879	13	CNS0148K
8	63.2	7.4	1042	13	CNS0148K
9	61.4	7.2	1001	13	CNS0155H
10	60.6	7.1	893	13	CNS013XE
11	60.2	7.1	1223	13	B12981
12	60	7.1	1204	13	CNS016E2

C 13	59.8	7.0	1101	13	CNS00210	AL061917 Drosophila
C 14	59.2	7.0	700	13	AO940248	AO940248 Sheared D
C 15	58.2	6.9	997	13	CNS005TE	AL060767 Drosophila
C 16	58	6.8	1101	13	CNS0006J	AL062049 Drosophila
C 17	57.8	6.8	1101	13	CNS001FB	AL060732 Drosophila
C 18	57.8	6.8	1125	10	AL547503	AL547503
C 19	57.6	6.8	1277	11	BF264952	BF264952 HY_CEA001
C 20	57.6	6.8	729	13	AO945618	AO945618 Sheared D
C 21	57.6	6.8	959	13	CNS00655	AL062806 Drosophila
C 22	57.6	6.8	1135	13	CNS0036Q	AL226115 Tetradon
C 23	57.6	6.8	1147	13	B13042	B13042 T30M24-Sp6.
C 24	57.4	6.8	804	13	B12681	B12681 F27D1-Sp6.1
C 25	57.4	6.8	897	13	CNS07ABZ	AL436389 T7 end of
C 26	57.4	6.8	938	13	CNS006RT	AL065906 Drosophila
C 27	57.4	6.8	1091	13	CNS014AC	AL063902 Drosophila
C 28	57.2	6.7	834	13	B12387	B12387 F21E20-Sp6.
C 29	57.2	6.7	964	13	CNS07EBR	AL441457 T7 end of
C 30	57.2	6.7	1101	13	CNS014XJ	AL104737 Drosophila
C 31	57	6.7	997	13	CNS0134P	AL102403 Drosophila
C 32	56.6	6.7	945	13	CNS04DOK	AL285149 Tetradon
C 33	56.4	6.6	942	13	CNS018GS	AL109318 Drosophila
C 34	56.4	6.6	974	13	CNS001TT	AL075432 Drosophila
C 35	56.4	6.6	1885	10	BE420745	BE420745 HMM002.B0
C 36	56.2	6.6	915	11	BG520365	BG520365 ps20C09.Y
C 37	56	6.6	907	13	CNS021J4	AL176953 Tetradon
C 38	55.8	6.6	895	13	CNS06SEV	AL396821 T7 end of
C 39	55.4	6.5	668	10	AL514901	AL514901
C 40	55.4	6.5	846	13	CNS00G25	AL072931 Drosophila
C 41	55.4	6.5	994	13	CNS04NOJ	AL298972 Tetradon
C 42	55.4	6.5	1101	13	CNS00088	AL063632 Drosophila
C 43	55.2	6.5	1034	13	CNS010B4	AL098746 Drosophila
C 44	55.2	6.5	1101	13	CNS008K2	AL057309 Drosophila
C 45	55	6.5	806	13	CNS04ABE	AL281759 Tetradon

ALIGNMENTS

RESULT 1	CNS06X9S	1007 bp	DNA	GSS	06-JUL-2001
LOCUS	T3 end of clone AX0A039F08 of library AX0A from strain CBS 7064				
DEFINITION	of Pichia farinosa, genomic survey sequence.				
ACCESSION	AL19462				
VERSION	AL19462.1	GI:12202640			
KEYWORDS	GSS.				
SOURCE	Pichia farinosa.				
ORGANISM	Pichia farinosa				
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.				
AUTHORS	1 (bases 1 to 1007) de Montigny,J., Spehner,C., Souciet,J., Tekala,F., Dujon,B., Wincker,P., Artiguenave,F. and Potier,S.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 15. Pichia sorbitophila				
JOURNAL	FEBS Lett. 487 (1), 87-90 (2000)				
MEDLINE	20584725				
REFERENCE	2 (bases 1 to 1007)				
AUTHORS	Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durans,P., Lepoint,A., Llorente,B., Malpertuy,A., Neugeilse,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies				
JOURNAL	FEBS Lett. 487 (1), 3-12 (2000)				
MEDLINE	20584711				
REFERENCE	3 (bases 1 to 1007)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremlieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :				

[illegible][illegible]

Oy	664	AAATTCGATGCAGCAACGAAGAAGCTTTTATAATTACAAGCTGGATCACAAG	723
Dd	862	ATMAATTAATAAAMAAAAATMAATTTMTTTTMMAMATATTAAMAMATMAAAMAAA	921
Oy	724	CATGCGTAGCAAGAGTGACGGAATCATTTAATCATACTGTGCACATATTAATGCT	783
Dd	922	AAAAAAAAATATAMAMMMWTATATTTTWTATTTAAATMTATMTWATTWATTMAMTWATATMT	981
Oy	784	AATCCCGATTGGAAGTACTGTTAAAGCTCAATTTAAAGCTTTAAACAGATATAA	840
Dd	982	TWTATWTATATATATWTATMTATATATTTAATAMMTATATTTAAMAAMATATATA	1038
RESULT	3		
LOCUS	AQ946120/c	GSS	27-JAN-2000
DEFINITION	Sheared DNA-46J23, TR Sheared DNA Trypanosoma brucei genomic clone		
ACCESSION	AQ946120		
VERSION	AQ946120.1	GI:6769385	
KEYWORDS	GSS.		
SOURCE	Trypanosoma brucei.		
ORGANISM	Trypanosoma brucei Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;		
REFERENCE	Trypanosoma. 1 (bases 1 to 641) El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C., Gerard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J., Fraser,C. and Adams,M. Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library Unpublished (1999)		
JOURNAL	Other-GSSS: Sheared DNA-46J23.TF		
COMMENT	Contact: Najib M. El-Sayed Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: nelsayed@tigr.org Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: http://www.tigr.org/tcdb/mdb/vbdb/ . Seq primer: MJ3-Reverse Class: Shotgun.		
FEATURES			
Source	Location/Qualifiers		
	1..641		
	/organism="Trypanosoma brucei"		
	/strain="TREU927/4 GUTat 10.1"		
	/db_xref="taxon:5691"		
	/clone="Sheared DNA-46J23"		
	/clone.lib="Sheared DNA"		
	/note="Vector: pUC18; Site.1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."		
BASE COUNT	213 a	35 c	10 g 383 t
ORIGIN			
Query Match	7.9%; Score 67; DB 13; Length 641;		
Best Local Similarity	44.7%; Pred. No. 0.0022;		
Matches	259; Conservative 0; Mismatches 320; Indels 0; Gaps 0;		
Oy	266	CTACGATACACTCTGGCAATTAATCAACGAATGTTACGCTTCATTAAGAAGTGACGGGGA	325

D	b		CTAGTAGTACGAGTAGTACTAGTACTACTATTAATAAATAATATATATATATAATA	578
O	y	326	CAAGTAGTGTTTCTATTATATAAAACGGAGATANGCTACCAGAGATACGACATGTAC	385
D	b	577	ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	518
O	y	386	GATGGTTTTTAATATTTAACAAATGAAAAAGTTATGTTCGAAGAATTTACTATAAAG	445
D	b	517	ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	458
O	y	446	ATCAGATTCAAGCGGCAGACAGTTAGTTTAAGCACATTTAAACATTTAGTGACAGTA	505
D	b	457	ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	398
O	y	506	CACATGCAATTTATTTAGTGGACAAAGTCGATTAAGTTTGAAAAGCCTTCCAG	565
D	b	397	ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	338
O	y	566	GTTCTAAATAACTGTTGATATAACGAGAACACACATTTGATGTAACTTCACAAAGCT	625
D	b	337	ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	278
O	y	626	ATGGTCATATAATAGTTTTCAATTAATCTACAAAACCAAATTTACGATGAACGAAA	685
D	b	277	ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	218
O	y	686	AAGAGTTGTTTAATATTTCCACAAGCTGTATCAAGACATGTGTAAGAGAGAATGAACG	745
D	b	217	ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	158
O	y	746	GGAATTCATTTAATCATCTGTGCCAATTTTAATGCTAAATCCGGTATTTGAAGTACTG	805
D	b	157	ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	98
O	y	806	TAAAGCGAATTTAAAGTTTAAACGATTTAAACGAGTAAGATA	844
D	b	97	ATAATAATAATAAGACGTGTTTGAGTGTGCTATTA	59

RESULT	4
CNS00396/c	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

CNS00396 1101 bp DNA GSS 03-JUN-1999
 BACR08Rk10 of RPC1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 AL063921
 AL063921.1 GI:4941778
 GSS.
 fruit fly.
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1101)
 Genoscope.
 Direct Submission
 submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 Evry cedex - FRANCE (E-mail : seque@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Oseegawa and
 Aaron Mammosser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPC1-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y2; cn bw sp. the same strain used for the BDGP's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1. 1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RBC1-98"
/clone="BAC08K10"
/note="end : TET3"
BASE COUNT 201 a 64 c 131 g 202 t 503 others
ORIGIN

Query Match 7.7%; Score 65.2; DB 13; Length 1101;

Best Local Similarity 14.4%; Pred. No. 0.0044;
Matches 75; Conservative 258; Mismatches 186; Indels 1; Gaps 1;

QY 330 TAGGTTTCTTCAATTAAGGAGATATGCTACCAAGAAAGATGACACATGATGATG 389
Db 1090 DDDTRKDDMDTKMTWKMDRADDRMAGDADRMAGDAGTMTATMTMTMTMTMTMT 1031

QY 390 GTTTTAAATATTAACAATGAAAAAGTTATGATGAAAGATATTACTATTAAGATCA 449
Db 1030 WMDKWWMTAAKTDPAWTMTWRADWAGRBGAKRDRDAATDADGAGRDRGRK-R 972

QY 450 GATTCAGGTGACAGCAGTGTAGATTAAAGCATTAAATGTCAGACGATACACA 509
Db 971 KDKRKRGDDDKGKGGKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 912

QY 510 TAGCATTATTATAGTGGCAAGATGCAATGATTTTGTAAAAAGCCTTCCAGGTTC 569
Db 911 DDDGKGRADDDTDTKDDDDKDDKDDKAGTGDATWMAAATDMMWMAADMMWMTW 852

QY 570 TAAATTAACCTGTGATTAACGAGAACAAATGATGATTAACATTCACAAGCTATG 629
Db 851 DAADDDWMAADRDMDAMWMDAMWMDAMWMDAMWMDAMWMDAMWMDAMWMDAMWMD 792

QY 630 GTCATATTAATGTTTCAATTAACATAAACAATTAAGCAATGACAGACAGCAAGA 689
Db 791 AADDDDAATTTTWTTTTTRDTDDKMTDTWTTRMAADDTWDRDDDDDRAGTGRKRR 732

QY 690 GTTGTATTAATTAATCAAGCTGTATCAAGCATGATGAAGAAAGAGAGAGAGAG 749
Db 731 RTWKRMRKRRDTRMDADADTDTRDRRRRGGDAGAGKGTGKRRRRDRAATWDRTDA 672

QY 750 ATCATTTAATCACTGTGCACATATTAATGCTAATGCGGCTATGGAAGTACTGTA 809
Db 671 WWDAAWMTTDTDTDDKRRRRRRRTTARAAMDWTWKAMDMAKMDKTRAD 612

QY 810 AGGTGAATTTAAAGTTTAAACAGATTAAGATACCAAG 849
Db 611 RWDRAADTWTADKARADWAKARARARARARARADR 572

RESULT 5
CNS0182P 1101 bp DNA GSS 26-Jul-1999
LOCUS Drosophila melanogaster genome survey sequence sp6 end of BAC
DEFINITION BACN37D10 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL108811
VERSION AL108811.1 GI:5629115
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Plasmid Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (22-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seque@genoscope.cns.fr

COMMENT - Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
<http://www.edgp.ebi.ac.uk> - . This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billard at CNRS (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Boucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES

source

1. 1101
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN37D10"
/note="end : SP6"
BASE COUNT 274 a 268 c 128 g 73 t 358 others
ORIGIN

Query Match 7.6%; Score 64.8; DB 13; Length 1101;

Best Local Similarity 22.2%; Pred. No. 0.0052;
Matches 105; Conservative 180; Mismatches 188; Indels 1; Gaps 1;

QY 366 AGAATATCGACACATGTGATGTTTAAATTTAAATTAAGTAAAGTATGATC 425
Db 585 DWMDRGAAGAAADADADADADADADADADADADADADADADADADADADADAD 644

QY 426 GAAATATTTACTAATTAAGATCAATTCAGATGACAGCAGTTTAAAGACAT 485
Db 645 AAAAGGKGTGKRAADGGGAAADAWRRGRGRGRRAAARAAADAKRAAARAA 704

QY 486 AAACATTAATGATGACAGTACATATGATATGATGACAAAGTCAATTAAT 544
Db 705 AAAAAAAGGKGRKRRKKGDKGKATTAAMAKGKMDGTATMTMTATADTAK 764

QY 545 ATTTGAAAGACCTTCCAGGTTGTAATTAATGATGATTAAGCAAGACAAATG 604
Db 765 AATDTRAKRAAAGRRRRRRAKTRARGRARRARRARRARRARRARRARRAD 824

QY 605 ATGTAACAATTCACAGGCTATGGTCAATTAATGATTTTCAATTAATCAAGCA 664
Db 825 KMDMAAAGGKRAAAGGKRAAAGGKRAAAGGKRAAAGGKRAAAGGKRAAAGG 884

QY 665 AAATTAAGCAATGACAGCAAGGATTTGTTAATTAATTAATTAATTAATTAAT 724
Db 885 RRRRRARRARRAADDTDKDMADATTDKTTTDDDDDDDDDDDDDDDDDDDDDD 944

QY 725 ATGTAAAGGAGAGTGAAGGGAATCATTAATCACTGTGCACATTAATTAATGCTA 784
Db 945 KWRDARADWATTAADDDGKMDKMGKRGKRGKMDKMDKMDKMDKMDKMDKMD 1004

QY 785 ATCCGCTATTAAGTACTGTAAGGATTAAGGATTAAGGATTAAGGATTAAGGATA 838
Db 1005 WWTTRKMDWMDGRGRGRMTTRKRWGAWRADAWRADDTDKTRTADKRR 1058

RESULT 6
CNS0022B 1059 bp DNA GSS 26-Jul-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION BACN01115 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL097133
VERSION AL097133.1 GI:5608744
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Plasmid Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1059)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (22-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seque@genoscope.cns.fr

AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : segeif@genoscope.cns.fr)

COMMENT
 - Web : www.genoscope.cns.fr
 - Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) -
 http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
 library (Dros BAC) was made by Alain Billard at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBelOBAC11.

FEATURES
 Location/Qualifiers
 1..1059
 /organism="Drosophila melanogaster"
 /plasmid="pBelOBAC11"
 /db_xref="taxon:7227"
 /clone_lib="DrosBAC"
 /clone="BACN01115"
 /note="end : 17"

BASE COUNT 14 a 21 c 66 g 760 t 198 others
 ORIGIN

Query Match 7.6%; Score 64.6; DB 13; Length 1059;
 Best Local Similarity 37.0%; Pred. No. 0.0058;
 Matches 312; Conservative 64; Mismatches 466; Indels 2; Gaps 1;

5 AATAAATGCAAAATATACAAATGTGACATGATTAAGTGCATGGCGACAGAGGTA 64
 843 AAAAAAAAAAAAAAAAAAADAADAAADKATADAAAABCBACACAAAAAAMACCKCA 784
 65 CAGTAAGATAGAGGTTTATGTAATAACAGTACATTAAGCTTTAAAGGTGAACAGTGG 124
 783 ACKBCKACAKKCCWAAATATAAAAMWCAKSKMBRACBCDMAAYAAACCKKMA 724
 125 GTCAAGCACTTTACACCAAGCGTGCAACAATTACATCATGATAAAGTAGAAAAAT 184
 723 BACKCAAAAKMNAAAAAAAAAAAAAAAAAAAKAKNKMKDKAKMNAAAAKKMAAA 664
 185 TAAGTAGTGTTCGGGATTTGCAGATTGAAGTACAGAGAAATTTAAGCCAAACA 244
 663 ADGAKAKAAATDKKKKKKAAKATGRDAKADDAKAKAAAAAADRGAA--AA 606
 245 ATACTTCAGATGACAAAGTACGATCAATCTGGGAATTAATCAACGAATGTTGCG 304
 605 AKAAGAGDAAGKAGGTGATGRDPAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 546
 305 TTCTATTAAGTGAAGCGGACAGAGTAGTCTTCTATTTATAAAGCGAGATATGCTAC 364
 545 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 486
 365 CAGAGATGACACATGATGATGTTTAAATTAACATGAAGAAAGTTATGTAT 424
 485 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 426
 425 CGAAGATTTACTATAAGGATCAGATTCAAGGTGACAGAGATTGATTTAAGCAGAT 484
 425 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 366
 485 TAAACATTATGTGACAGTACACATGATTTATATGTGACAAAGTCAATTGCG 544
 365 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 306
 545 ATTGTAAGAAAGCTTTCCAGGTTCTAATTAATCTGTATATACAGAGACACATTTG 604
 305 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 246
 605 ATGTAAACAATTCACAAGGCTATGGCTCATATATAGTTTTCATTAATTAACAACA 664
 245 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 186

QY 665 AATTACGATGACAGCAAAAGACTTTGTTATATATTCACAGAGTGTGATCAAGAC 724
 DB 185 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 126
 QY 725 ATGTAAGCAAGAGTGAACGGGAATCATTTAATCATCTGTGACAAATATTATGCTA 784
 DB 125 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 66
 QY 785 ATGCCGTTTGAAGTACTGTAAAGTGAATTAAGTTTAAACGAGTAAAGATA 844
 DB 65 AANNNAAAAAAAAAAAAAAAAAAAVANAAGAAAGNMAAAAAAAAAATCTTAANA 6
 QY 845 CCAA 848
 DB 5 CTTAA 2

RESULT 7
 LOCUS CNS01JRG/c 879 bp DNA GSS 12-JUN-2001
 DEFINITION Anopheles gambiae GSS t7 end of clone 14D07 of NotreDame1 library
 from strain PEST of Anopheles gambiae (African malaria mosquito),
 genomic survey sequence.
 ACCESSION AL147405.1 GI:7005551
 VERSION AL147405
 KEYWORDS GSS.
 SOURCE African malaria mosquito.
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
 Culicoidae; Anopheles.
 1 (bases 1 to 879)
 Genoscope.
 DIRECT SUBMISSION
 Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : segeif@genoscope.cns.fr)
 - Web : www.genoscope.cns.fr
 2 (bases 1 to 879)
 Roth,C.W., Brey,P.T., Ke.Z., Collins,F.H. and Weissenbach,J.
 Direct Submission
 Submitted (16-FEB-2000) BMH, Institut Pasteur, 25, rue du Dr.
 Roux, Paris 75015, France
 This clone is from an A. gambiae BAC library provided by F.H.
 Collins and sequenced by Genoscope in collaboration with the
 Laboratory of Biochem. and Biol. Molec. of Insects, Institut
 Pasteur.

FEATURES
 Location/Qualifiers
 1..879
 /organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="14D07"
 /clone_lib="Notredame1"
 /note="end : 17"

BASE COUNT 66 a 38 c 1 g 670 t 104 others
 ORIGIN

Query Match 7.5%; Score 64; DB 13; Length 879;
 Best Local Similarity 38.9%; Pred. No. 0.0077;
 Matches 271; Conservative 50; Mismatches 375; Indels 1; Gaps 1;

QY 153 AACATTTACATTCATGATTAAGTACAAAAATTAAGTATGTTTCGGGATTCAGAAATT 212
 DB 843 AAAAAAAAAAAAAAAAAAAAAAAAAAGTATATATATTTGAAATGWTATTAATAAAAAA 784
 QY 213 TGAAGTACAAGGAAGAAATTTAAGCAACAATTAATCTTCAGATGACAAAGTACTACGAT 272
 DB 783 AAAAAAAGAAAAAAGCTTAAATTAATAAATAAATAAATATGTGAATAAATAAATAATATAGTAA 724
 QY 273 AACATCTGGAAATTAATCAACGAATGTACGTTCAATAAAGTGAAGCGGACAAAGTAG 332
 DB 723 AATATGAGMAMDKRAKMAAAAGKTBRAKRGMAAAAAAAAAAAAAAAAAATATAAAGKMA 664

QY	333	T-GTTTCTATTATTAATGCGGAGGATTTATGACACGAAGATATACGACATGACATGCT	391
Db	663	TAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	604
QY	392	TTTAAATATTTAACAATGAAAAAGTTATGATCGAAAGATATTTACTATTAAGATCGA	451
Db	603	CGWGTDTDAATAAATATGAAAAAAGAGACGKRGARAAAGAAKAKAKAAAAA	544
QY	452	TTCAAGGTGACACGACGTTTAAACACATTTAAACATTAATGTGACAGTACACATA	511
Db	543	AAAAATGTTAAATAATGAAAAAATAKATATRAAAAAAAAAAATAGAAAKRGAGAAAAAR	484
QY	512	GCAATTTATTTATGTCGACAAAGTCATTTACTGATTTTGAAGAAAGCTTTCCAGTGTGA	571
Db	483	AGATWAKAAAAAATAAATAATDAAAAATKAWTTTBAAAAAAGAKAAGRTATTTGDR	424
QY	572	AAATPACTGTTGATPANTCGAAGAACACAAATTTGATGTAAACATTCACAGGCTATGCGT	631
Db	423	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	364
QY	632	CATATTAATAGTTTTCATTTAATCAATTAACCAATTAATTCGAAATTAACGCAAAAAAGT	691
Db	363	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	304
QY	692	TTGTTAATTAATTCACAAAGCTTGTATCAGAGCATGTTAGGAAAGTGAAGCGGAAT	751
Db	303	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	244
QY	752	CATTTAATCATATCTGTGACAAATATTAATGCTAATGCCGATTTGAAAGTACTGTTAAG	811
Db	243	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	184
QY	812	GTGAATTTAAAGTTTAAACAGCATTAAGATACCA	848
Db	183	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATA	147
RESULT	8		
CNS0148K/c			
LOCUS	CNS0148K	1042 bp	DNA
DEFINITION	Drosophila melanogaster genome survey sequence spe end of BAC		
	BACN11E16 of DrosBAC library from Drosophila melanogaster (fruit		
	fly), genomic survey sequence.		
ACCESSION	AL103838		
VERSION	ALI03838.1	GI:5615449	
KEYWORDS	GSS.		
SOURCE	fruit fly.		
ORGANISM	Plasmod Drosophila melanogaster		
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	1 (bases 1 to 1042)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequençage :		
	BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr		
COMMENT	- Web : www.genoscope.cns.fr)		
	Determination of this BAC-end sequence was carried out as part of a		
	collaboration with the European Drosophila Genome Project (EDGP) -		
	http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC		
	library (DrosBAC) was made by Alain Billaud at CEPH (Centre		
	d'Etude du Polymorphisme Humain) with funding provided by a MRC		
	project grant. The DNA was prepared from embryos by Alain Bucheton		
	and Genevieve Payan. It has been constructed in the vector		
	pbel0BAC11.		

```
FEATURES
source
location/qualifiers
1..1042
/organism="Drosophila melanogaster"
/plasmid="pBelosK11"
/db_xref="taxon:7227"
/clone_id="DrosBAC"
/clone="BAC111E16"
```

BASE COUNT	68 a	55 c	41 g	690 t	188 others
ORIGIN					
Query Match	7.4%: Score 63.2; DB 13; Length 1042;				
Best Local Similarity	35.0%: Pred. No. 0.011;				
Matches 295; Conservative	91;	Mismatches 455;	Indels 1;	Gaps 1;	

Query Match	Similarity	35:08	Pred. No. 0.011	DB 13	Length 1042
Best Local	Conservative	91	Mismatches 455	Indels 1	Gaps
QY	8	AAAAATGGAAAAATCAAAATAGTGCATCGATTGATTAAGAGCGATGGCCGACAGCGGTACAG	67		
Db	1012	AMNAAVAAVAAASRAASMAAAAGSKMTAAHASTAGSMASCSCTGMMCCAAAAGAAVRR	953		
QY	68	TAAAGATGAGCGGTTATATGTTAAACAGTACCATTAACTGTTAAAGGTGAACAGGTGGTC	127		
Db	952	RRARRRRRRRRARABGGAARRRRTAARNTMAAAGAACAMAMWNTDSAAAGAAAGSVH	893		
QY	128	AAGCGATTATACCCAGACGCGTCCACAAATTCATTCATGATTAAGGTGAAGAAATTTAA	187		
Db	892	TMAAARAAMATMKAAAMKACATAMTSASAMWNTSAAGTGATMTAMVTAAVSCMCMW	833		
QY	188	GTGATGTTTCGGGATTTTCGAAATTTTAAGTACAGAGAGAAATTTTAACGCAACAATA	247		
Db	832	YAAATTSAMGAAASMGGAAMWTHHCCATTTHARGMAAMGAAAGSACHTAABYTABIT	773		
QY	248	CTTAGATGACAAAGTACGTACGTAAACATCTGGGATTAATCAAGATGTTACGGTTC	307		
Db	772	MMATATASCATATWAMSAKCAAAAAMASSTSRCSMMMAAMTATGSMAAAATGSTATAT	713		
QY	308	ATAAAGTGAACGGGAGCAACATGTGTCTTTCTATTATAACGGGAGATATGCTACAC	367		
Db	712	ACACCHAAAGSTTWAMAMCMCKGKATATATCATCTGCACSCATGACATATCGATGTWTMA	653		
QY	368	AAGA-TACGACACATGTACGATGCTTTTAAATTTTAAACATGAAAAAGTATGTATCG	426		
Db	652	ACGAGTGGGGAATTTTACATGTGRTHTWTAAAGACATTTTGAAAAAAGAAAAAAG	593		
QY	427	AAAGATTTTACTATTAAGATCAGATTCAAGGTGACAGCATTTAGTTTAACACATTA	486		
Db	592	AAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG	533		
QY	487	AACATTAATGTGACAGGTACACATAGCAATTTATATAGTGGACAAAGTGAATTTACGAT	546		
Db	532	AAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG	473		
QY	547	TTTGAAAAAGCCTTCCAGGTTCTAAATTAACGTGTGATATACGAAGACACATTTGAT	606		
Db	472	AAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG	413		
QY	607	GTAACAATTCACAAGGCTATGGTCAATATATATAGTTTTCATTAATCTACAAACCAA	666		
Db	412	AAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG	353		
QY	667	ATTACGATTAAGACGCAAAAAAGATTGTTAATATATTCACAAGCTTGTATCAAGGAT	726		
Db	352	AAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG	293		
QY	727	GGTAGAGACAGATGAGACGGGAATCATTTAATCATCTGTGCACATATTAAATGTAAAT	786		
Db	292	GAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA	233		
QY	787	GCCGGATTGGAAGTACTGTAAAAAGTGAATTAAAGTTTAAACAGATTAAGATACC	846		
Db	232	AAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG	173		
QY	847	AA 848			
Db	172	AA 171			

RESULT	9
CNS0155H	
LOCUS	CNS0155H 1001 bp DNA GSS 26-JUL-1999

DEFINITION	Drosophila melanogaster genome survey sequence sp6 end of BAC BACNI3C23 of drosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
ACCESSION	AL105023.1 GI:5617037			
VERSION	GSS.			
KEYWORDS	fruit fly.			
SOURCE	Plasmod Drosophila melanogaster			
-ORGANISM	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
REFERENCE	1 (bases 1 to 1001)			
AUTHORS	Genoscope.			
TITLE	Direct Submission			
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)			
COMMENT	determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.			
FEATURES	Location/Qualifiers			
source	1..1001			
	/organism="Drosophila melanogaster"			
	/plasmid="pBelOBAC11"			
	/db_xref="taxon:7227"			
	/clone_lib="DrosBAC"			
	/clone="BACNI3C23"			
	/note="end : SP6"			
BASE COUNT	266 a 219 c 134 g 150 t 232 others			
ORIGIN				
Query Match	7.2%; Score 61.4; DB 13; Length 1001;			
Best Local Similarity	34.2%; Pred. No. 0.024;			
Matches 163; Conservative	87; Mismatches 219; Indels 7; Gaps 1;			
QY	214	GAAGTACAAGGAAATTTAAACGCAACAATATCTTCAGATGACAAAGTAGCTACGATA	273	
Db	525	GAGADARAATGAGTAAAWRAAGCAAAACGRVRAWMDGAGADAGATATNAGAA	584	
QY	274	ACATCTGGGATTAATATCAGCAATGTTCGGTCTTAATAAGTGAAGCGGAAACAAGTGT	333	
Db	585	AAAKRGMAAABDAGASAAARRTKKGARGGAAAAAGMDADDAATWMKRAARA	644	
QY	334	GTTTTCCTTTTAAACGGAGATGTGCTACAGAAAGTAGACACATGTACGATGTTT	393	
Db	645	DADMAAAACAAAHHVHAANAARAKARRRAAAAAARAGCAAA--	697	
QY	394	TTAAATATTAACAATGAAAAAGTTATGTATCGAAAGATATTACTATAAGATCAGATT	453	
Db	698	GRAAARDBRAAAAAATAAAMWAATRAFAAARDBRAAAAAATTAAMWTAAMWMAAMW	757	
QY	454	CAAGTGCAGCGCACTTGATTAAAGCATTAACATTAATGTGACAGGTACACATAGC	513	
Db	758	WWAATAATAAADDTTWTATWAATAAATAAAAWMTAAAAAATATTAAWMAWMAAAATWMA	817	
QY	514	AATATATTAAGTGCACAAGTGCATTTATCTGATTTTGAAAAAGCCTTCCACAGTCTAAA	573	
Db	818	AAATTTWMAAAWMTAAWMAAAAAAATTTTAATTTTTTAAAMAAWMAATWMAATITAAAA	877	
QY	574	ATTAAGTGTGATTAATAGCAAGAACAAATGTATGATACAAATTTCCAAAGCTATGGTCA	633	
Db	878	DAAAAAAAMAAWMAATTAATTTAAATTTTAAWMTTAAWMAADTTTWTTTTAAWMAAAA	937	
QY	634	TATTAATCTTTTTCATTAATCTACAAAACCAAAATTAAGATGAACACAGCAAAAAA	689	
Db	938	TAAWMTAAWMTAAWMTAAWMTAAWMTAAWMAAAWMAWMAWMTAAWMTAAWMAAAA	993	

RESULT 10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						</
-----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

QY	579	TGTTGAAATATCGAAGAACCAATTTGATGACCAATCCCAAGGCGTATGTCATTAA	638
Db	520	TATTTAANNAATATAAAATTTAAAAATTTMTTAAATTTT-----TTTAATTTAATATAA	5722
QY	639	TAGTTTTCATTTACTACAAACCAAAATTTAGCATGTAAGAACAGCAAAAAGATTGTAA	658
Db	573	AATATATATTATAAAAAATAAAATTTAAAAAATAAAATTTAAATTTGAATTTTTH	6322
QY	699	TATATCAACAGCTGTGATACAGCATGGTAAGGAAGTAAGTGAACGGGAATCATTTAA	758
Db	633	TTTMAAATATAAAARWTGGAAARGGGCGGGKTTTGGATGRRRRGATATATAAAATATAA	6522
QY	759	TCATACTGTGCACAATATTAATGCTAATGCCGATATTTGAAGTACTGTAAAGGTGAAAT	818
Db	693	AATATATATTTTTTTATTTTTRATTTTWTMAAAATTTAAATTTTTTTTATKRAAARADMAATK	7522
QY	819	AAAAGTTTAA	830
Db	753	RATRTKTKTKDM	764

RESULT	11
LOCUS	B12981/C
DEFINITION	B12981 1223 bp DNA GSS 14-MAY-1997
ACCESSION	T24D11-Sp6 TAWO Arabidopsis thaliana genomic clone T24D11, DNA sequence.
VERSION	B12981
KEYWORDS	B12981.1 GI:2094103
SOURCE	GSS.
ORGANISM	thale cress. Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 1223) Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and Peng,J.
TITLE	BAC End Sequences at ATGC
JOURNAL	Unpublished (1997)
COMMENT	Contact: Ecker J.

Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@atgenome.bio.upenn.edu
Seq primer: Sp6
Class: BAC ends
High quality sequence start: 199
High quality sequence stop: 271.

FEATURES	
Source	Location/Qualifiers
1. .1223	
	/organism="Arabidopsis thaliana"
	/strain="Columbia"
	/db_xref="taxon:3702"
	/clone="T24D11"
	/clone_1id="TAMU"
	/sex="hermaphrodite"
	/note="Vector: BelobAC11; Site_1: HindIII; Site_2: HindIII"
	; Produced by Rod Wing"
BASE COUNT	70 a 50 c 53 g 678 t 372 others
RRIGIN	

Query Match	7.1%	Score	60.2	DB	13	Length	1223
Best Local Similarity	33.9%	Pred. No.	0.039				
Matches	236	Conservative	0	Mismatches	460	Indels	0
						Gaps	0

[illegible]

OY	213	TGAATTCACAGGAAATAATTATACCACAATAACTCTCAGTGGCAAGTGTCATCGAT	272
Dd	1087	NNNNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAN	1028
OY	273	AACATCTGGGAATAAATCAAGCAATGTTCAGGTTCAAAGTGAAGCGGACAAGTAG	332
Dd	1027	NNAANNANAANNAAAAAAANNNNNNTATTTAANNAAAAAANANNAANNAANAAA	968
OY	333	TGTTTTCTATTATAAACGGGAGATATGCTACCAAGAATGCAGACATGTCCGATGGTT	392
Dd	967	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANNNNNNNAN	908
OY	393	TTTAAATATTAACAATGAAAAAGTTATGTATCGAAGATATTACTATPAAGATCATGAT	452
Dd	907	NNNNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANANNNANNNANNAANAAAAA	848
OY	453	TCAAGTGAGACAGCAGTTAGATTTAAGCACATTAACACTTAATGTGACAGGTACATAG	512
Dd	847	NAAAAAAAAAAAAAAAAAANNAANNAANNAANNAAAAAAAAAAANANNAANAAAAAN	788
OY	513	CAATTATTATAGTGACAAAGTGCATTACTGATTTTGAAAAAGCCTTCCAGGTTCTAA	572
Dd	787	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANNNNNANNNANNNANAA	728
OY	573	AATAACTGTGTATAATACGAGAACACACATTGATGTAACAATTCACAAAGCTATGGGTC	632
Dd	727	AAAAANNNNNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAN	668
OY	633	ATATPATAGTTTTCATTAACACTCAAAACCAAAATTACGAATGAACGCCAAAAAGATT	692
Dd	667	ANNNNAAAAAAAAANANNAANNAANNAANNAANNAANNAANNAANNAANNAAN	608
OY	693	TGTTAATATTCACACAGCTGGTATCCAAGCATGGTAAAGGAAGAAGTGAACGGGAATTC	752
Dd	607	NNNNNAAAAAAAAAAAAAAAAAAAAAAAAANNAANNNANNNANNNANNNANNAANNA	548
OY	753	ATTTAATCATCTGTGCAAAATATTAAATGCTAATGCCGATATTGAAGGTACTGTAAAAG	812
Dd	547	ANNAANNAANNNANNAANNAANNAANNAANNAANNAANNAANNAANNAANNNNN	488
OY	813	TGAATTAAGAAGTTTAAACAAGATRAAAGATRCGA	848
Dd	487	NNNANNNNNNNNNNNAAAAAAAAAAAAAAAAANAA	452

RESULT	12
CNS016E2/c	
LOCUS	CNS016E2 1204 bp DNA GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC BACN15A12 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL106628
VERSION	AL106628.1 GI:5622852
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Plasmid Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1204)
REFERENCE	Genoscope.
AUTHORS	Direct Submission
TITLE	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL	

COMMENT
Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> . This *Drosophila melanogaster* BAC library (Dros BAC) was made by Alain Billaut at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pReloBAC11.


```

VERSION      AQ940248.1  GI:6763513
KEYWORDS     GSS.
SOURCE       Trypanosoma brucei.
ORGANISM     Trypanosoma brucei
              Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;
              Trypanosoma.
REFERENCE    1 (bases 1 to 700)
AUTHORS      El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
              Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Doneison,J.,
              Fraser,C., and Adams,M.
TITLE        Determination of clone end sequences from Trypanosoma brucei GUTat
              10.1 sheared DNA library
JOURNAL      Unpublished (1999)
COMMENT      Other_GSSs: Sheared DNA-42E21.TR
              Contact: Najib M. El-Sayed
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: nelsayed@tigr.org
              Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
              DNA library constructed at TIGR. Clones will be available for
              distribution through ATCC. Sheared DNA end sequences search page:
              http://www.tigr.org/cdb/mdb/cdbd/.
              Seq primer: M13-Forward
              Class: Shotgun.
FEATURES
  source      1..700
              Location/Qualifiers
                /organism="Trypanosoma brucei"
                /strain="TREU927/4 GUTat 10.1"
                /db_xref="taxon:3691"
                /clone_lib="Sheared DNA-42E21"
                /clone_1lb="Sheared DNA"
                /note="Vector: pUC18; Site1: SmaI; Constructed at The
                Institute for Genomic Research (TIGR), Rockville, MD.
                Genomic DNA isolated from a cloned population of
                Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
                sheared to give a tight size distribution (approx 2 kb).
                The v + i method used for the library construction is
                described in detail in Smith, H.O. and Venter, J.C.
                (Making small insert libraries for whole genome shotgun
                sequencing projects. In Genome Sequencing: A Practical
                Approach, eds. M. Vaudin and B. Bartell, Oxford University
                Press, 1999)."
BASE COUNT   374 a 43 c 41 g 242 t
ORIGIN
Query Match 7.0%; Score 59.2; DB 13; Length 700;
Best Local Similarity 46.6%; Pred. No. 0.065;
Matches 190; Conservative 0; Mismatches 218; Indels 0; Gaps 0;
QY 397 AATATTAACAATAAAGTATGATCGAAGATATTCATAAAGTCAGATTCAA 456
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 290 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 349
QY 457 GGTGACAGCAGTATGATTAAACATTAAACATTAAATGACAGTACATAGCAAT 516
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 350 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 409
QY 517 TATTAATGAGGACAAAGTCATCTGATTTGAAAAGCCCTTCAGGTTCTAAATA 576
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 410 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 469
QY 577 ACAGTGAATATGGAAGAACATGATGATGACATTCACAGGCTTGGGTCTAT 636
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 470 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 529
QY 637 AATAGTTTTTCAATTAATCAAAACAAATTAAGCAATGACAGCAAAAGATTGTT 696
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 530 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 589

```

```

QY 697 AATTAATTCACAGCTTGATTCAGACATGTAAGAGAACTGAACGGGAATCATTT 756
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 590 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 649
QY 757 AATCATTACGTGCACATATTAATGCTAATGCCGATGATGAGCTACT 804
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 650 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 697
RESULT 15
LOCUS       CNS005TE/C
DEFINITION  CNS005TE. 997 bp DNA GSS 03-JUN-1999
            Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL060767
VERSION     AL060767.1 GI:4943573
KEYWORDS    GSS.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 997)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
            Web : www.genoscope.cns.fr)
FEATURES
  source      1..997
              Location/Qualifiers
                /organism="Drosophila melanogaster"
                /db_xref="taxon:7227"
                /clone_1lb="RPCI-98"
                /clone="BACR12K22"
                /note="end : TET3"
BASE COUNT   89 a 99 c 13 g 258 t 538 others
ORIGIN
Query Match 6.9%; Score 58.2; DB 13; Length 997;
Best Local Similarity 18.0%; Pred. No. 0.095;
Matches 129; Conservative 173; Mismatches 415; Indels 0; Gaps 0;
QY 4 GATTAATATGGAATAATACAAATGCTGACATGATTAAGTGCATGCGGACAGCGT 63
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 898 RRAAARAAADARARARARARARARARARARARARARARARARARARARAR 839
QY 64 ACAGTGAATATGGAAGAACATGATGATGACATTCATTAAGGTTGAACAGGTG 123
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 838 AGRGRCGRGRRRARRRRARARARARARARARARARARARARARARARARAR 779
QY 124 GGTCAACCACTTATTCACACGAGCGTCAACAATTACATTCATGATAAGTAGAAAA 183
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 778 GRRRGCGRRRRRRARRARARARARARARARARARARARARARARARARAR 719
QY 184 TTAAGTATGTTTCGGGATTTGCAATTTGAAGTCAAGAAATTTTAACGCAACA 243
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```
Db 718 GARRRRRRRRAAAGAGAAARRRRRRRGAGARRRRRGRRRGAGARRRRRRGRMR 659
QY 244 AATTACTTCAGATGACAAGTAAGTACGATACATCGGAGATAATCAACGATGTTACG 303
Db 658 TTRARRRRRRARRAGARRRRRRRGRRRRRRRGRRRRRRRGRRRRRRRGRRRRRGA 599
QY 304 GTTCATAAAGTGAACGGGACCAAGTACTGTTTCTATTATAAAACGGAGATATGCTA 363
Db 598 RRGRRRAGRRRRRRARRAGARRRRRRRGRRRRRRRGRRRRRRRGRRRRRRRGRRRRRGA 539
QY 364 CCAGAAGATACACACATGTCAGATGTTTAAATATTAAACATGAAAAAGTTATGTA 423
Db 538 GRAGARRRRRRRRRRRRRRRRRRRRRGARRRRRRRGRRRRRRRRRRRRRAAARAR 479
QY 424 TCGAAGATATTACTATAAGATCAGATTCAGGTGACAGCAGTTAGATTAAACACA 483
Db 478 RRARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 419
QY 484 TTAACATTAATGTGACAGTACACATAGCAATTATTATAGTGACAAAGTCAATTACT 543
Db 418 NNNNNNNNCATANNNTNTTTTNTNTNTNANNNTTTTNTNTNNNTTNTTN 359
QY 544 GATTTGAAAAAGCCTTCCAGGTCTAAATPACTGTTGATPATAGAGAACAACAAATT 603
Db 358 NNTTANTNTNTNNNTTNNATTTGACCTTCTATATTATTATTAATTAAACMTNN 299
QY 604 GATGTAACAATCCACAAGGCTATGGGTCATATATAGTTTCAATTACTACAAACC 663
Db 298 NNNNAANNNTANANAAAAANNANNNNNNNNNAAAAATANANNNNNNNNNNNAAAAATA 239
QY 664 AAAATTACGAATGACAGCAAAAGAGTTGTAATPATTCACAGCCTGGTATCAA 720
Db 238 AAAANNAANNNNNNNNNANANANNNNNNNNNAAAAATAHMSNNANYARABMBBA 182
```

Search completed: January 29, 2002, 22:42:29
Job time: 4603 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 23:21:02 ; Search time 56.04 Seconds
(without alignments) 286.810 Million cell updates/sec

Title: US-09-813-820-4

Sequence: 1 MRGSHNNHNSDQKVAATIT.....GIEGVKGEIKVLKQDKDK 211

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	911	81.2	1185	2 A42404	collagen adhesin -
2	194	17.3	37	2 A48620	adhesin - Staphylo
3	131	11.7	1092	2 S42798	fibronectin-bindin
4	119	10.6	336	2 D86710	collagen adhesin f
5	100	8.9	1536	2 A43855	high-molecular-we
6	99.5	8.9	352	2 A81888	probable surface f
7	98.5	8.8	4919	2 T31105	hypothetical prote
8	98	8.7	1117	2 S33851	fibronectin-bindin
9	96.5	8.6	848	2 C70203	DNA topoisomerase
10	96	8.6	392	2 A53580	neurexin III beta
11	96	8.6	704	1 S39004	finger protein MSN
12	96	8.5	359	2 E86706	hypothetical prote
13	95	8.5	666	2 A42296	lysozyme 2 (EC 3.2
14	95	8.5	2020	2 C48339	ABC-type transport
15	94.5	8.4	2902	2 C71953	toxin-like outer m
16	94	8.4	2380	2 E71604	hypothetical prote
17	93.5	8.3	532	2 T06029	hypothetical prote
18	93.5	8.3	1297	2 S39791	neurotoxin - Clost
19	93.5	8.3	2340	2 B71704	cell surface antig
20	92.5	8.2	711	2 S73898	DNA topoisomerase
21	92.5	8.2	1943	2 B64596	toxin-like outer m
22	92.5	8.2	2910	2 T28156	DNA-directed RNA p
23	92	8.2	1037	2 T13350	transcription fact
24	92	8.2	1983	2 G86643	hypothetical prote
25	92	8.2	2500	2 S16619	hypothetical prote
26	91.5	8.2	261	2 S16619	opacity protein op
27	91.5	8.2	2150	2 S71629	sensory transducti
28	91	8.1	463	2 A44808	cellulase (EC 3.2.
29	91	8.1	2269	2 T28677	riophly protein -

30	90.5	8.1	784	2 JH0101	apolipoprotein B-1
31	90	8.0	669	2 S14535	asparagine-rich pr
32	90	8.0	858	1 A42239	adenylate cyclase
33	90	8.0	4688	2 F82885	hypothetical prote
34	89.5	8.0	416	1 A41267	transcription fact
35	89.5	8.0	454	2 T26296	hypothetical prote
36	89.5	8.0	496	2 S61327	IGA-specific metal
37	89.5	8.0	1584	2 T22674	hypothetical prote
38	89.5	8.0	1815	2 C81169	IGA-specific metal
39	89.5	8.0	2893	2 A64556	toxin-like outer m
40	89	7.9	570	2 T46261	hypothetical prote
41	89	7.9	599	2 S68118	laccase (EC 1.10.3
42	89	7.9	834	2 B82940	conserved hypothet
43	89	7.9	1166	2 T28680	fibrirogen-binding
44	89	7.9	1449	2 T30552	glucosyltransferas
45	89	7.9	1553	2 T18502	hypothetical prote

ALIGNMENTS

RESULT 1
A42404
collagen adhesin - Staphylococcus aureus
C:Species: Staphylococcus aureus
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 10-Nov-1995
C:Accession: A42404; S27665
R:Patil, J.M.; Jonsson, H.; Guss, B.; Switalski, L.M.; Wiberg, K.; Lindberg, M.; Hook
J. Biol. Chem. 267, 4766-4772, 1992
A:Title: Molecular characterization and expression of a gene encoding a Staphylococcus
A:Reference number: A42404; MUID:92165839
A:Contents: FDA 574
A:Accession: A42404
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1185 <PAT>
A:Cross-references: EMBL:M81736
A>Note: sequence extracted from NCBI backbone (NCBIP:83982)

Query Match 81.2%; Score 911; DB 2; Length 1185;
Best Local Similarity 90.1%; Pred. No. 4.3e-58;
Matches 182; Conservative 2; Mismatches 16; Indels 2; Gaps 2;
QY 12 SDDVAVTTSGNKSNTVTHKSEAGTSVFY-RTGDM-LPEDTTHVRFPLINNEKSYV 69
Db 144 SDDKVAATTSNGKSTNVYGIWKREPVYFLINKSGKICQEDTTHVRFPLINNEKSYV 203
QY 70 SKDTTKKQIOGGQOLDSTININVTGHSNMYSGQSAITDFEKAFFPSKTIYDNTKMTI 129
Db 204 SKDTTKKQIOGGQOLDSTININVTGHSNMYSGQSAITDFEKAFFPSKTIYDNTKMTI 263
QY 130 DVTIPQGGYSNFSINIKTKITNEQKFEYNNNSQAWQEHGKEEVNKSFNHTVHNINA 189
Db 264 DVTIPQGGYSNFSINIKTKITNEQKFEYNNNSQAWQEHGKEEVNKSFNHTVHNINA 323
QY 190 NAGIEGVKGEIKVLKQDKDK 211
Db 324 NAGIEGVKGEIKVLKQDKDK 345
RESULT 2
A48620
adhesin - Staphylococcus aureus (fragment)
C:Species: Staphylococcus aureus
C:Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 03-Mar-1995
C:Accession: A48620
R:Patil, J.M.; Boles, J.O.; Hook, M.
Biochemistry 32, 11428-11435, 1993
A:Title: Identification and biochemical characterization of the ligand binding domain
A:Reference number: A48620; MUID:94032261
A:Contents: FDA 574
A:Accession: A48620

A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid; protein
A:Residues: 1-37 <PAT>
A:Note: sequence extracted from NCBI backbone (NCBIP:138726)

Query Match	17.3%;	Score 194;	DB 2;	Length 37;
Best Local Similarity	100.0%;	Pred. No. 3.8e-08;		
Matches	37;	Conservative	0;	Mismatches 0;
				Indels 0;
				Gaps 0;
QY	19	ITSGNKSTNVVHKSEAGTSSVFYFKTGDMLPEDTTH	55	
Db	1	ITSGNKSTNVVHKSEAGTSSVFYFKTGDMLPEDTTH	37	

RESULT 3
S42798
fibronectin-binding protein - Streptococcus "equisimilis"
C:Species: Streptococcus "equisimilis"
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 15-Oct-1999
C:Accession: S42798
R:Signaes, C.
submitted to the EMBL Data Library, December 1993
A:Reference number: S42798
A:Accession: S42798
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1092 <Sig>
A:Cross-references: NID:q436135; PIDD:CAA82330.1; PID:q436136
C:Keywords: fibronectin binding

Query Match	11.7%	Score 131;	DB 2;	Length 1092;
Best Local Similarity	26.5%	Pred. No.	0.097;	
Matches	58;	Conservative	31;	Mismatches 88;
				Indels 42;
				Gaps 13

QY 19 ITSGKSNVNVHSEACTS--VFYKTKGDMLEPETH---VRWPLNIN-NEKSVASMDI 73
 Db 152 VISGKHAHEVVIQKRGSTARDNITYKSGDI--PTQMEDSVRMSFTFNARKSTNGYGF 209
 QY 74 TIKDQIOGGQOLDL---SPLININVTGTHSNVYSGOASITDFE----KAPPGSKITVD 123
 Db 210 LVTFDLIDSTMTMVDYKKNLREKVAINFETGG---WIGDELLSKNKGMLSKAEANYGIVAE 265
 QY 124 NTKNTIDVTIQQ---GIGSY-----NSFSINYTKTKITNEOQKEPVNNSQAWY-OEHGKE 173
 Db 266 FSGGTGSVINIPEKIQDYDNGSYQOELNESEMOIHLVAKI---KKEVLENSSTIEYVNESKV 321
 QY 174 EVNGKSF-----NHT---VHNINANAGIEGTAKGELKVLK 205
 Db 322 EYRGENMPIDPNSISAYVQILRGGMAKGIYVAGEVRIILK 360

RESULT 4
D86710
collagen adhesin [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
C:Accession: D86710
R:Biological, A.; Winkler, P.; Manger, S.; Jallou, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. In press, 2001
A:Title: The complete genome sequence of the lactic acid bacterium.
A:Reference number: A86625
A:Accession: D86710
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-336 <SNO>
A:Cross-references: GB:AE005176; NID:912723592; PIDN:AAK04782.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
C:Gene: lacC

Query Match	10.6%;	Score 119;	DB 2;	Length 336;
Best Local Similarity	25.2%;	Pred. No. 0.16;		
Matches	52;	Conservative	28;	Mismatches 68;
				Indels 58;
				Gaps 11;

[illegible]

RESULT 5
A43855
high-molecular-weight surface-exposed protein HMW1 - Haemophilus influenzae
C:Species: Haemophilus influenzae
C:Date: 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1995
C:Accession: A43855
R:Barenkamp, S.J., Leininger, E.
Infected. Immun. 60, 1302-1313, 1992
A:Title: Cloning, expression, and DNA sequence analysis of genes encoding nontypeable
detella pertussis.
A:Reference number: A43855; MUID:92192797
A:Accession: A43855
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1536 <BAR>
A:Cross-references: GB:008876; GB:M84616; NTD:g475770; PIDD:AAA20527.1; PTD:g475771
A:Note: Sequence inconsistent with the nucleotide translation
A:Note: Sequence extracted from NCBI database (NCBIIN:89235; NCBIIP:89239)

[illegible]

RESULT 6
A81888
probable surface fibril protein NMA1200 [imported] - Neisseria meningitidis (strain Z
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: A81888
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
R.; Holroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491
;Reference number: A81775; MUID:20222536


```

Db      514 ARKLNNTLPTTIGAAVINLLEKTFPVLELNTSNMEEKIDIAIGKLDKIKYLSKFN 573
OY      104 GOSAITD-----FEKAFPGSKITVNTKNTIDVTIPQG-YGSY-----NSFS 144
Db      574 GKGLKLDVWOLEPKTIDSSFEFTVIESQKT---ENKNSITVITINICKYGPYLFFKGNYS 630
OY      145 INKTKITNEOQK-----FVNSQAVYQOEKKEVNGKS--FNHTYH-NI-----NANAG 192
Db      631 INAKTLEMLYKKDETEKINELEKPNILGVDPLGLNVIFFKNTIYGNITVQLGEPTHAP 630
OY      193 IEGTVKSG---ELKVKODK 208
Db      691 QETTKGKPKKIKITAKK 709

```

```

RESULT 10
A53580
neurexin III beta precursor (clone PB794-7) - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 12-Apr-1995 #sequence_revision 12-Apr-1995 #text_change 17-Mar-1999
C:Accession: A53580
R:Ushakov, Y.A.; Hata, Y.; Ichchenko, K.; Moosaw, C.; Afendis, S.; Slaughter, C.A.; S
J. Biol. Chem. 269, 11987-11992, 1994
A:Title: Conserved domain structure of beta-neurexins. Unusual cleaved signal sequences
A:Reference number: A53580; MUID:94216308
A:Accession: A53580
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-392 <USH>
A:Cross-references: GB:L27864
C:Keywords: alternative splicing

```

```

Query Match      8.6%; Score 96; DB 2; Length 392;
Best local Similarity 19.6%; Pred. No. 9.1;
Matches 47; Conservative 35; Mismatches 74; Indels 84; Gaps 10;

```

```

OY      5 HHHHHGSDDKVATITSGKSTNVYVHKS-----EAGTSVYVYKGTGMDLPEDTTHVR 57
Db      52 HEHHFGSK-----HHSVPISIRSPVSLRGHAGATYFFKSGGLI-----YT 96
OY      58 WFLN-----INNEKSYSKDITTDQI-----QG--GQQLDLSLNI 92
Db      97 WPAHDRPSTRSDMLAVGFSTYVDGILVRIDSAFGDLQHLIEQKIGVNVNIGTVDI 156
OY      93 NTGTSNTSYSGSATIDFEKAPFGSKITVDNTKNTIDVTIPQGYGSYNSFSIN--YKTK 150
Db      157 STKEETPVNDGKYHVVRFRNGNATLOYD-----WPNVNHPTG 198
OY      151 ITNEOQKEFVNS-----QAWYQEHGKEEVNGKSFNHTVHNINANAGIEGTVGEL 201
Db      199 NNDNEFRQVWVKQIPKYNRPVEEMLQEKGRQ-----LTIFNQAOALIGKRGKRL 250

```

```

RESULT 11
S39004
finger protein MSN2 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YM9532.O2C; protein YMR037C
C:Species: Saccharomyces cerevisiae
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C:Accession: S39004; S52886; A48131
R:Estreich, F.; Carlson, M.
Mol. Cell. Biol. 13, 3872-3881, 1993
A:Title: Two homologous zinc finger genes identified by multiplicity suppression in a SNF1
A:Reference number: A48131; MUID:93309420
A:Accession: S39004
A:Molecule type: DNA
A:Residues: 1-704 <EST>
A:Cross-references: EMBL:L08838; NID:g349594; PIDN:AAA34806.1; PID:g349595
R:Odell, C.; Bowman, S.
Submitted to the EMBL Data Library, February 1995
A:Reference number: S52885

```

```

A:Accession: S52886
A:Molecule type: DNA
A:Residues: 1-704 <CODE>
A:Cross-references: EMBL:248502; NID:g695715; PIDN:CAA8403.1; PID:g695717; GSPDB:GNO
C:Genetics:
A:Gene: SGD:MSN2; MIPS:YMR037C; MIPS:YMR037C
A:Cross-references: SGD:S0004640; MIPS:YMR037C
A:Map position: 13R
C:Superfamily: finger protein MSN2
C:Keywords: DNA binding; nucleus; transcription regulation; zinc finger
F:649-665/Region: zinc finger CCH motif
F:678-698/Region: zinc finger CCH motif

```

```

Query Match      8.6%; Score 96; DB 1; Length 704;
Best local Similarity 25.4%; Pred. No. 19;
Matches 64; Conservative 31; Mismatches 97; Indels 60; Gaps 13;

```

```

OY      12 SDRKVAITTSGNKSTNVYVHKSAGTSVYVYKGTGMDLPEDTTHVRKFLINNEKSYVSK 71
Db      229 SDNYSYNSISNSNSNSTGNLN---SYFNSINIDSMLDYVSDLLNDDDDDTNLSR 284
OY      72 ---DITIKDIOG-----GQQLD-----STLNINVTGHSNYSGOSAIT 109
Db      285 RRRSDV--ITNQFSPMTNRSNSISHSLDMNHPKINPSNRNTNLNIT--TNSTSSMASPNT 342
OY      110 DFEKAPFGSKITVDNTRN---TIDVTIPQGYGSYN-----SFSI 145
Db      343 TTMNANADSNI--AGNPKNDATIDNELTQILNRYNNMNFNDNLGTSIGKKNKSPSSD 401
OY      146 NYTKITNEOQ-KEPVNSO--AWYQEHGKEEVNGKSFNHTVHNINANAG---IEGTVG 199
Db      402 NAMTKINPQOQOOLNRYVQHKQLTSSHNNSSTNKSFSNDLSRRQRAFLPIIDSLSY 461
OY      200 ELKVKODKTK 211
Db      462 DL-VNKKDDEPK 472

```

```

RESULT 12
E86706
hypothetical protein ygea (imported) - Lactococcus lactis subsp. lactis (strain IL140
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
C:Accession: E86706
R:Boletín, A.; Winkler, P.; Manger, S.; Jallón, O.; Malarme, K.; Welssenbach, J.; Eh
Genome Res. in press, 2001
A:Title: The complete genome sequence of the lactic acid bacterium.
A:Reference number: A86625
A:Accession: E86706
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-359 <STO>
A:Cross-references: GB:AE005176; NID:g12723558; PIDN:AAK04751.1; GSPDB:GNO0146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: ygea

```

```

Query Match      8.5%; Score 95; DB 2; Length 359;
Best local Similarity 21.7%; Pred. No. 9.6;
Matches 58; Conservative 36; Mismatches 89; Indels 84; Gaps 12;

```

```

OY      19 ITSGNKS---TNTVTKSEAGTSVYVYKGTGMDLPEDTTHVRKFLINNEKSYVSDIR 75
Db      101 LTYMNSPDTYSVFKKIDDDKNVVGKTFEKKVTDRFETS--INADEISF-----FTS 154
OY      76 KQDIOGQQLDLSLNI-----NVGTGTH---SNYSGOSAITDFEKAAPGSKITVDNTR 126
Db      155 KIDAEIKKGDIVTIVKRRPNILGLVYIMSGVQSKKNINOHKS--NNESNSNNTK 212
OY      127 -----NTIDVTIPQGYGSYNSFSINYKTKITNEOQKEFV 160

```

Db 213 IYNLHNTFTSPSNWVKQLTNSNPDNTINFLLDQVLYNHYEFETN-ESIFNNRYNSFL 271
Qy 161 -----NN-----SQAWY--OEHGKEVNGK-----SFNHT 183
Db 272 SSKLNTSNMMSDFTATNETEGSSSESYAWYISLKNKEKNGEMVLLVDYGGISFTIY 331
Qy 184 VHNINAMAGIEGTVKGLKVLKODKDT 210
Db 332 VSGDKFSKNTGEPDNNLLDSIKYTKRT 358

RESULT 13
A42296
lys2zyme 2 (EC 3.2.1.-) precursor - Enterococcus hirae (ATCC 9790)
C:Species: Enterococcus hirae
C:Date: 10-Jul-1992 #sequence_revision 18-Sep-1992 #text_change 15-Oct-1999
C:Accession: A42296
R:Chu, C.P.; Kariyama, R.; Daneo-Moore, L.; Shockman, G.D.
J. Bacteriol. 174, 1619-1625, 1992
A:Title: Cloning and sequence analysis of the muramidase-2 gene from Enterococcus hirae.
A:Reference number: A42296; MUID:92165737
A:Accession: A42296
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-666 <CHU>
A:Cross-references: GB:M77639; NID:q148303; PIDN:AAA24776.1; PID:q148304
A:Note: the authors translated the codon GGT for residue 171 as Tyr
C:Keywords: glycosidase; hydrolase

Query Match 8.5%; Score 95; DB 2; Length 666;
Best Local Similarity 24.0%; Pred. No. 21;
Matches 49; Conservative 28; Mismatches 81; Indels 46; Gaps 11;

Qy 20 TSGKSTNVVHKSEAGTSSVFYKTKDMLPEDTTH-----VWFLLINNEKSYVS 70
Db 242 TSGMSSGSAAT-----TGTTYTKSGDSV-WGISHFGITMAQLIER-NKIKNFIYPG 293
Qy 71 KDTIKQIQGGQQLDSTLNI-----NVTG-THSNYSQGSATIDFEKAPGSKITV 122
Db 294 QKLIK-----GGQSAGSSTNTGNNASSGNTSGNTSGTSGQAT-----GAKYTV 340
Qy 123 DNFNTKIDVTIPQGYSPSINYKTKITNEOQKEFVNNNSQAWYOEHGKEVNGKSPNH 182
Db 341 KSGDSVKRIANDHC-----ISKMLIE-WNNIKNFYVPEQLVSKSSASGTSNT 393
Qy 183 TVHNINAMAGIEG-TVKGLKVLK 205
Db 394 STGNTSSNTANTGSTTSGSYTVK 417

RESULT 14
C48399
ABC-type transport protein ydbA.2 - Escherichia coli
C:Species: Escherichia coli
C:Date: 19-Nov-1993 #sequence_revision 16-Oct-1998 #text_change 08-Oct-1999
C:Accession: C48399; D64891; H64891
R:Moszer, I.; Glaser, P.; Danchin, A.
Biochimie 73, 1361-1374, 1991
A:Title: Multiple IS insertion sequences near the replication terminus in Escherichia coli
A:Reference number: A48399; MUID:92190338
A:Accession: C48399
A:Molecule type: DNA
A:Residues: 464-2020 <MOS>
A:Cross-references: GB:D85081; NID:q3041754
A:Experimental source: strain K-12
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBI:88089, NCBI:88090)
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co-
A:Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617

A:Accession: D64891
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-839, 'IDLPLYFQTSVIT' <BLA1>
A:Cross-references: GB:AE000237; GB:U00096; NID:q1787665; PIDN:AACT4483.1; PID:q17876
A:Experimental source: strain K-12, substrain M61655
A:Accession: H64891
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'M', 915-2020 <BLA2>
A:Cross-references: GB:AE000237; GB:U00096; NID:q1787665; PIDN:AACT4487.1; PID:q17876
A:Experimental source: strain K-12, substrain M61655
C:Genetics:
A:Gene: ydbA_2
A:Start codon: GTG

Query Match 8.5%; Score 95; DB 2; Length 2020;
Best Local Similarity 25.1%; Pred. No. 84;
Matches 69; Conservative 30; Mismatches 92; Indels 84; Gaps 16;

Qy 5 HHHHNG-----SDDKATITSGNK-----STNVTVHKSSEA 35
Db 94 HHHNNNSPLRPTPPDDSDDRPPTPGDEILIPDDPDPTPPKPVSFNNVDILDKTEK 153
Qy 36 GTS---SVFY---KTGDMLEDTTHVWFINT-----NNEKST--VSKDITTKQIQG 81
Db 154 TLTITRDSVFYTFENADGTISLQDSNGRKATINLQIDRANTVALLEGVSADGATKMQYNH 213
Qy 82 GQQL-----DSTLNI-----VNGTSHNYSQGSATIDFEK-----AFPSK 119
Db 214 NGELVITGDNATVNNNGKTYVDGKDSGTETENGNGK-VIDGGLDVSGGCHDITDS 272
Qy 120 ITVNTKNTIDVTIPQGYG---SYNSFSINY--KTKITNEOQKEFVNNNSQAWYOEHGKEE 174
Db 273 ATVDN-KGTMTVTDPESMGIDDDKATVNEGESTINGTGQINGDDATANNNGKTT 331
Qy 175 VNGKSFNHTVHNINAMAGIEGTVKGLKVLKODK 209
Db 332 VDKRDSGTG--EINGNG-----KVI-QDDG 354

RESULT 15
C71953
toxIn-like outer membrane protein jhp0274 - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: C71953
R:Alm, R.A.; Ling, L.S.L.; Molt, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.
; Ives, C.; Gibson, R.; Metberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
A:Reference number: A71800; MUID:99120557
A:Accession: C71953
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2902 <ARN>
A:Cross-references: GB:AE001464; GB:AE001439; NID:q4154789; PIDN:AD05855.1; PID:q415
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp0274

Query Match 8.4%; Score 94.5; DB 2; Length 2902;
Best Local Similarity 23.1%; Pred. No. 1.4e+02;
Matches 50; Conservative 26; Mismatches 63; Indels 77; Gaps 11;

Qy 23 NKSNTNVVHKSEAGTSSVFYKTKDMLPEDTTHVWFINTNNKSYSKDITIKDQ----- 78
Db 1335 NLSQNSAINSNHSTLEL-----QGLDLNDITS-----LNLQSAINSNNAITINDYASLI 1386
Qy 79 IQGGQQLD-----LSTLNI-----NVGTSHNYSQGSATID- 110

```
Db 1387 ASNGSHLNENGAVNENSANITTSLSSSIVFKGAVSLRGOFNLNNSSLDFOGSSAITSN 1446
QY 111 -----PEKAPGSKITVDNWKNTIDYTIPOGYGS-----YNSFSINYKTKITNEQOKEFY 160
Db 1447 TAFNFYDNAFSOSPITF---HQALDIKVPISLGCNLLNPNNSSVLNK----- 1491
QY 161 NNSQAWYQEHGK-----EEVNGKSFNHTVHNT 187
Db 1492 -NSQVFSDOGSLNLIANIIDLSDLNGN--KNRYYN 1524
```

Search completed: January 30, 2002, 00:03:33
Job time: 2551 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 22:44:32 ; Search time 67.56 Seconds
(without alignments)
231.342 Million cell updates/sec

Title: US-09-813-820-4

Perfect score: 1122
Sequence: 1 MRGSHHHHSGDDKVAITIT.....GIEGVKGEKLVKQDKDPTK 211

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A.GeneSeq_1101:*

- 1: /SIDSR/gcgdata/geneSeq/AA1980.DAT:*
- 2: /SIDSR/gcgdata/geneSeq/AA1981.DAT:*
- 3: /SIDSR/gcgdata/geneSeq/AA1982.DAT:*
- 4: /SIDSR/gcgdata/geneSeq/AA1983.DAT:*
- 5: /SIDSR/gcgdata/geneSeq/AA1984.DAT:*
- 6: /SIDSR/gcgdata/geneSeq/AA1985.DAT:*
- 7: /SIDSR/gcgdata/geneSeq/AA1986.DAT:*
- 8: /SIDSR/gcgdata/geneSeq/AA1987.DAT:*
- 9: /SIDSR/gcgdata/geneSeq/AA1988.DAT:*
- 10: /SIDSR/gcgdata/geneSeq/AA1989.DAT:*
- 11: /SIDSR/gcgdata/geneSeq/AA1990.DAT:*
- 12: /SIDSR/gcgdata/geneSeq/AA1991.DAT:*
- 13: /SIDSR/gcgdata/geneSeq/AA1992.DAT:*
- 14: /SIDSR/gcgdata/geneSeq/AA1993.DAT:*
- 15: /SIDSR/gcgdata/geneSeq/AA1994.DAT:*
- 16: /SIDSR/gcgdata/geneSeq/AA1995.DAT:*
- 17: /SIDSR/gcgdata/geneSeq/AA1996.DAT:*
- 18: /SIDSR/gcgdata/geneSeq/AA1997.DAT:*
- 19: /SIDSR/gcgdata/geneSeq/AA1998.DAT:*
- 20: /SIDSR/gcgdata/geneSeq/AA1999.DAT:*
- 21: /SIDSR/gcgdata/geneSeq/AA2000.DAT:*
- 22: /SIDSR/gcgdata/geneSeq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1122	100.0	211	AAW31553	Collagen binding p
2	1054.5	94.0	512	AAW31554	Collagen binding p
3	904	80.6	1185	AAW31555	Collagen binding p
4	830	74.0	159	AAW31552	Collagen binding p
5	-178	15.9	458	AAW31557	Collagen binding A
6	165.5	14.8	21	AAW31558	Collagen binding A
7	125.5	11.2	345	AAW31555	Collagen binding A
8	116	10.3	473	AAW31555	Collagen binding A
9	114.5	10.2	688	AAW31555	Collagen binding A
10	114.5	10.2	2032	AAW31555	Collagen binding A
11	114.5	10.2	2032	AAW31555	Collagen binding A

12	114.5	10.2	2032	AAW0242	Enterococcus faeca
13	113.5	10.1	184	AAW32917	E.coli optimised K
14	113.5	10.1	184	AAW10309	Human KGF-2 constr
15	113.5	10.1	184	AAW1673	PQ60-Cys37 constr
16	110.5	9.8	1338	AAW1731	High molecular wtl
17	110.5	9.8	1598	AAW30291	Non-typeable Haemo
18	110	9.8	493	AAW30992	P. falciparum spor
19	110	9.8	2314	AAW69136	M. catarrhalis les
20	109	9.7	1536	AAW30283	Non-typeable Haemo
21	105.5	9.4	342	AAW39498	Human apoAIV mutel
22	104.5	9.3	1529	AAW1732	High molecular wtl
23	103.5	9.2	139	AAW31556	Fibronectin-bindin
24	101.5	9.0	847	AAW06400	Class A starch bra
25	101.5	9.0	1601	AAW30292	Non-typeable Haemo
26	100	8.9	537	AAW60452	Sequence of the As
27	100	8.9	1095	AAW01847	Haemophilus influe
28	100	8.9	1536	AAW1723	High molecular wtl
29	100	8.9	1536	AAW1725	High molecular wtl
30	100	8.9	1536	AAW63505	Haemophilus influe
31	100	8.9	1536	AAW01846	Haemophilus influe
32	99.5	8.9	389	AAW23745	A surface protein
33	99.5	8.9	589	AAW06173	N. meningitidis p2
34	99.5	8.9	592	AAW06180	Human apoAIV mutel
35	98.5	8.8	346	AAW39493	Fibroblast growth
36	98.5	8.8	351	AAW56000	His rrf10 protein
37	97.5	8.7	195	AAW59053	Nucleotide sequenc
38	97	8.6	437	AAW37738	Clostridium botull
39	97	8.6	451	AAW68398	S. epidermidis ope
40	96.5	8.6	1155	AAW82343	Cell wall protein
41	96.5	8.6	1802	AAW83170	Staph. epidermidis
42	96.5	8.6	329	AAW70119	Human recombinant
43	96	8.6	667	AAW44778	Human membrane-ass
44	96	8.6	667	AAW29654	Enterococcus hirae
45	95	8.5	666	AAW85291	

ALIGNMENTS

RESULT	ID	AAW31553	standard; Protein; 211 AA.
XX	XX	AAW31553:	
XX	XX	21-MAY-1998	(first entry)
XX	XX	Collagen binding protein M31	epitope.
XX	XX	Collagen binding protein; cna gene; sepsis; infection;	
XX	XX	Microbial surface component recognising adhesive matrix molecule;	
XX	XX	MSCRAMM; adhesin; vaccine; immunisation; diagnosis; therapy;	
XX	XX	epitope M31.	
XX	XX	Staphylococcus aureus.	
XX	XX	Key	Location/Qualifiers
XX	XX	Peptide	1..12
XX	XX	Protein	/note="vector POE30-derived peptide"
XX	XX		13..211
XX	XX		/note="epitope M31"
XX	XX	W09743314-A2.	
XX	XX	20-NOV-1997.	
XX	XX	14-MAY-1997;	97MO-US08210.
XX	XX	16-MAY-1996;	96US-0017678.
XX	XX	(UABR-) UAB RES FOUND.	
XX	XX	(TEXA) UNIV TEXAS A & M SYSTEM.	

PI Hook M, House-Pompeo K, Patti JM, Sthanam N, Symersky J;
 XX WPI; 1998-008801/01.
 DR N-PSDB; AAT93437.
 XX
 PT Antibody that interacts with collagen binding domain of
 PT Staphylococcal cna gene product - useful to prevent bacterial sepsis
 PT in animal infected with Staphylococcus aureus
 XX
 PS Claim 31; Page 115-116; 143pp; English.
 XX
 CC This protein comprises Staphylococcus aureus collagen binding
 CC protein (CBP) epitope M31, i.e. amino acids 61-343 of full-length
 CC CBP, plus a vector-derived N-terminal peptide. Claimed 441, 849
 CC and 1500 bp nucleic acid sequences (see AAT93436-38) respectively
 CC encode CBP epitopes M17, M31 and M55 (see AAW31552-54) that confer
 CC protection against S. aureus infection. These nucleic acid
 CC sequences can be used in the recombinant production of the CBP
 CC epitopes. The CBP protein and antigenic epitopes are contemplated
 CC for use in the treatment of pathological infections, especially to
 CC prevent bacterial adhesion to collagen. The claimed nucleic acids
 CC as well as claimed anti-CBP antibodies will also be of use in
 CC screening, diagnostic and therapeutic applications including active
 CC and passive immunisation and methods for the prevention of
 CC bacterial colonisation in an animal such as a human. The CBP
 CC epitopes are also contemplated for use in the preparation of
 CC vaccines and as carrier proteins in vaccine formulations, as well
 CC as in the formulation of compositions for the prevention of S.
 CC aureus infection.
 XX
 SQ Sequence 211 AA;

Query Match 100.0%; Score 1122; DB 19; Length 211;
 Best Local Similarity 100.0%; Pred. No. 1.2e-86;
 Matches 211; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRGSHHHHHGSDKYATTTGSKNSTNVYHKSEAGTSSVFYKKGDMLEPDTTHVRWFL 60
 DB 1 mrgshhhhhgsdkaqltsqkscnvtvkhseagtsvfyykkgdmlpedtthvrwfl 60
 QY 61 NINNESYVSKDTIRDOIGGQDLSTLNINVTGTHSNVYSGSATIDPEKAFPGSKTI 120
 DB 61 ninnesyvskdltlkdgq199gqldstlninvgtshnyysgsatldfeka1pyski 120
 QY 121 TYDNTKNTIDVTIPQGYGSYNSFSINVKTKITNEOQKEPVNNSOAMQOEKGEVNGKSF 180
 DB 121 tydntkntldvtlpggygsynsfisnykktikneqkfevmnsqawygelngeevngksf 180
 QY 181 NHTVHNINANAGIEGTGKGLKVLKODKDTK 211
 DB 181 nhtvhnnanaagiegvtgkglkvlkgdkdtk 211

RESULT 2
 AAW31554
 ID AAW31554 standard; Protein; 512 AA.
 XX
 AC AAW31554;
 XX
 DT 21-MAY-1998 (first entry)
 XX
 DE Collagen binding protein M55 epitope.
 XX
 KW Collagen binding protein; cna gene; sepsis; infection;
 KW microbial surface component regonising adhesive matrix molecule;
 KW MSCRAMM; adhesin; vaccine; immunisation; diagnosis; therapy;
 KW epitope M55.
 XX
 OS staphylococcus aureus.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..12

FT FT /note="vector pQE30-derived peptide"
 FT Protein 13..512
 FT /note="epitope M17"
 XX
 XX
 PN WO9743314-A2.
 XX
 PD 20-NOV-1997.
 XX
 PF 14-MAY-1997; 97WO-US08210.
 XX
 PR 16-MAY-1996; 96US-0017678.
 XX
 PA (UABR-) UAB RES FOUND.
 PA (TEXA) UNTV TEXAS A & M SYSTEM.
 PI Hook M, House-Pompeo K, Patti JM, Sthanam N, Symersky J;
 DR WPI; 1998-008801/01.
 DR N-PSDB; AAT93438.
 XX
 XX
 PT Antibody that interacts with collagen binding domain of
 PT Staphylococcal cna gene product - useful to prevent bacterial sepsis
 PT in animal infected with Staphylococcus aureus
 XX
 PS Claim 31; Page 117-119; 143pp; English.
 XX
 CC This protein comprises Staphylococcus aureus collagen binding
 CC protein (CBP) epitope M55, i.e. amino acids 30-531 of full-length
 CC CBP, plus a vector-derived N-terminal peptide. Claimed 441, 849
 CC and 1500 bp nucleic acid sequences (see AAT93436-38) respectively
 CC encode CBP epitopes M17, M31 and M55 (see AAW31552-54) that confer
 CC protection against S. aureus infection. These nucleic acid
 CC sequences can be used in the recombinant production of the CBP
 CC epitopes. The CBP protein and antigenic epitopes are contemplated
 CC for use in the treatment of pathological infections, especially to
 CC prevent bacterial adhesion to collagen. The claimed nucleic acids
 CC as well as claimed anti-CBP antibodies will also be of use in
 CC screening, diagnostic and therapeutic applications including active
 CC and passive immunisation and methods for the prevention of
 CC bacterial colonisation in an animal such as a human. The CBP
 CC epitopes are also contemplated for use in the preparation of
 CC vaccines and as carrier proteins in vaccine formulations, as well
 CC as in the formulation of compositions for the prevention of S.
 CC aureus infection.
 XX
 SQ Sequence 512 AA;

Query Match 94.0%; Score 1054.5; DB 19; Length 512;
 Best Local Similarity 64.7%; Pred. No. 1.8e-80;
 Matches 211; Conservative 0; Mismatches 0; Indels 115; Gaps 1;

QY 1 MRGSHHHHHG----- 11
 DB 1 mrgshhhhhgsardlstlnvtalvpskiedgkltvkmfdknkg1qngdmikvwaw 60
 QY 12 ----- 11
 DB 61 p1sgtvlkiegsktvp1tvkqgvgqav1tpdgat1itfndkvekl1dsvsgfaefevgrn 120
 QY 12 -----SDKRVATITSGNKSSTNVYHKSEAGTSSVFYKKGDMLEPDTTHVRWFLINNE 65
 DB 121 l1gtntsdskvalitsgnkscnvtvkhseagtsvfyykkgdmlpedtthvrwfl1nne 180
 QY 66 KSYVSKDITRKQDIOGGQDLSTLNINVTGTHSNVYSGSATIDPEKAFPGSKTIYDNT 125
 DB 181 ksyvskdltlkdgq199gqldstlninvgtshnyysgsatldfeka1pysk1tvdnt 240
 QY 126 KNTIDVTIPQGYGSYNSFSINVKTKITNEOQKEPVNNSOAMQOEKGEVNGKSFNHTVA 185
 DB 241 kn1tdvtlpggygsynsfisnykktikneqkfevmnsqawygelngeevngksfnhtvh 300
 QY 186 NINANAGIEGTGKGLKVLKODKDTK 211

Db 301 nlnanagiegtyvkgelkvkqdkdk 326

RESULT 3
AAR22675
ID AAR22675 standard; Protein; 1185 AA.

XX AAR22675;

AC 04-DEC-1992 (first entry)

XX Collagen binding protein.

DE CBP; collagen binding protein; mastitis; arthritis.

XX Staphylococcus aureus.

XX Key Location/Qualifiers

FT Region 30..1185 /note="claim 11; page 29-30"

FT Peptide 1..29 /label="sig_peptide"

FT Region 30..534 /label="A"

FT /note="see CC"

FT Region 535..721 /label="B1"

FT /note="see CC"

FT Region 722..908 /label="B2"

FT /note="see CC"

FT Region 909..1095 /label="B3"

FT /note="see CC"

FT Region 1096..1159 /label="W"

FT /note="see CC"

FT Region 1160..1179 /label="M"

FT /note="see CC"

FT Region 1180..1185 /note="charged C-terminal"

FT Region

XX WO9207002-A.

XX 30-APR-1992.

XX 22-OCT-1991; 91WO-SE00207.

XX 22-OCT-1990; 90SE-0003374.

XX (ALFA) ALFA LAVAL AGRI INT AB.

XX Guss BM, Hook M, Jonsson H, Lindberg KM, Patti J, Signaes LC;

XX Switajski LM;

XX WPI; 1992-167099/20.

XX N-PSDB; AA024123.

XX Hybrid DNA molecule encoding S.aureus collagen binding protein -

XX protein is expressed in E.coli and used for diagnosis e.g. of

XX septic arthritis

XX Disclosure; Fig 2; 40pp; English.

CC to mediate the binding of the protein to the cell wall. The amino

CC acid sequence nearest to the C-terminal end consists of a long

CC stretch of hydrophobic residues followed by some charged amino acids

CC This region is called M.

CC The CBP can be used for immunisation pref. in combination with a

CC fusion protein, e.g. for vaccination of ruminants against mastitis

CC caused by staphylococcal infections. It can also be used to block

CC infection in an open skin wound, e.g. for blocking protein receptors

CC or by immunisation. In the latter, the host produces specific

CC antibodies which block the adherence of the bacterial strains to

CC damaged tissue. This treatment can be used for septic arthritis

CC and tissue damage of e.g. skin, connective tissue, and mucous

CC membranes. Dosage for immunisation is 0.5-5 microg CBP/Kg; for

CC topical admin. the protein is used at a concn. of 25-250 microg/ml.

XX Sequence 1185 AA;

Query Match 80.6%; Score 904; DB 13; Length 1185;
Best Local Similarity 89.6%; Pred. No. 2,4e-67;
Matches 180; Conservative 3; Mismatches 16; Indels 2; Gaps 2;
OY 13 DDKVVATTSGNKTNTVHRSKSEAGTSVFY-RTGDM-LPEDTHVAFININNEKSYVS 70
DB 145 ddkvatltsygnkstnvygkvrkrepvflnksqkcygedtlhvflninnexyvs 204
OY 71 KDITIKDQIQGGQQLDISTLININVTGHSNYSQSAITDFEKAFPKSKITVDNFKNTID 130
DB 205 kdlitkdelqggqqldistlnlvtgchsnysqsaltdfekafrskltvdntkntid 264
OY 131 VTTPQGYGSYNSFSINKTKITNEQKEFYNNNSQAWQHEGKEVNNKSFNHTVHNINAN 190
DB 265 vtppqgygsynsfisinktkitneqgketyvnsqawyghegkeevngksfnhtvhninan 324
OY 191 AGIEGTWKGELVKLKDPRK 211
DB 325 aglegtyvkgelkvkqdkdk 345

RESULT 4

AAW31552
ID AAW31552 standard; Protein; 159 AA.

XX AAW31552;

XX 21-MAY-1998 (first entry)

XX Collagen binding protein M17 epitope.

XX Collagen binding protein; cna gene; sepsis; infection;

XX microbial surface component regionising adhesive matrix molecule;

XX MSCRAMM; adhesin; vaccine; immunisation; diagnosis; therapy;

XX epitope M17.

XX Staphylococcus aureus.

XX Key Location/Qualifiers

FT Peptide 1..12 /note="vector pQE30-derived peptide"

FT Protein 13..159 /note="epitope M17"

XX WO9743314-A2.

XX 20-NOV-1997.

XX 14-MAY-1997; 97WO-US08210.

XX 16-MAY-1996; 96US-0017678.

XX (UABR-) UAB RES FOUND.

XX (TEXA) UNIV TEXAS A & M SYSTEM.

PI Hook M, House-Pompeo K, Patil JM, Sthanam N, Symersky J;
 XX WPI: 1998-008601/01.
 DR N-PSDB: AAT93436.
 XX
 XX Antibody that interacts with collagen binding domain of
 PT Staphylococcal cna gene product - useful to prevent bacterial sepsis
 PI in animal infected with Staphylococcus aureus
 XX
 XX Claim 31; Page 114; 143pp; English.
 PS
 XX This protein comprises Staphylococcus aureus collagen binding
 CC protein (CBP) epitope M17, i.e. amino acids 151-297 of full-length
 CC CBP, plus a vector-derived N-terminal peptide. Claimed 441, 849
 CC and 1500 bp nucleic acid sequences (see AAT93436-38) respectively
 CC encode CBP epitopes M17, M31 and M55 (see AAW31552-54) that confer
 CC protection against S. aureus infection. These nucleic acid
 CC sequences can be used in the recombinant production of the CBP
 CC epitopes. The CBP protein and antigenic epitopes are contemplated
 CC for use in the treatment of pathological infections, especially to
 CC prevent bacterial adhesion to collagen. The claimed nucleic acids
 CC as well as claimed anti-CBP antibodies will also be of use in
 CC screening, diagnostic and therapeutic applications including active
 CC and passive immunisation and methods for the prevention of
 CC bacterial colonisation in an animal such as a human. The CBP
 CC epitopes are also contemplated for use in the preparation of
 CC vaccines and as carrier proteins in vaccine formulations, as well
 CC as in the formulation of compositions for the prevention of S.
 CC aureus infection.
 CC
 XX Sequence 159 AA:
 SQ
 Query Match 74.0%; Score 830; DB 19; Length 159;
 Best Local Similarity 96.4%; Pred. No. 2.7e-62;
 Matches 159; Conservative 0; Mismatches 0; Indels 6; Gaps 1;
 QY 1 MRSQHNNHSGDDKAYATTSNGKSTNTVHKSEAGTSSVFYKTDMLPEDTHVWFL 60
 Db 1 mrsqhnnhsgd-----ltsqnsklnvvhkseagtsvfyktdmlpedtlhvwfl 54
 QY 61 NNNKSYVSKDITTKDOIGGQGLDSTLNTNVTGTHSNYVSGSALIDPEKAFPGSK 120
 Db 55 nnnksyvsksdltkdgqggqgldstlnnvtgthsnnyvsgsalidpekafrpgsk 114
 QY 121 FVDNFKNTIDVTIPQGYGSNSFSINYSKTKITNEQKKEFVNNSQA 165
 Db 115 tvdnfkntidvtipqgygsnfsinsyktkitneqkkelvnmsqa 159

RESULT 5
 AAB37667
 ID AAB37667 standard; protein; 458 AA.
 XX
 AC AAB37667;
 XX
 DT 28-FEB-2001 (first entry)
 XX
 DE Collagen binding Ace protein.
 XX
 KM Collagen binding Ace protein; antibacterial; immunostimulant; vaccine;
 KW extracellular matrix-binding protein; lamin binding; bacterial infection.
 XX
 OS Enterococcus faecalis.
 XX
 PN WO200068242-A1.
 PD 16-NOV-2000.
 XX
 PF 10-MAY-2000; 2000WO-US12590.
 XX
 PR 10-MAY-1999; 99US-0133334.
 XX

PA (TEXA) UNIV TEXAS A & M SYSTEM.
 PA (TEXA) UNIV TEXAS MEDICAL SCHOOL.
 XX
 XX Rich RL, Kriekemeyer B, Owens RT, Hook M, Murray BE;
 PI Nallapareddy SR, Qin X, Weinstein GM, Singh KV, Duh R;
 XX WPI: 2000-687639/67.
 DR N-PSDB: AAC83828.
 XX
 XX New collagen-binding protein from Enterococcus, useful e.g. in
 PT protective vaccines, for diagnosis and treatment of Enterococcal
 PT infections and for screening for compounds that inhibit collagen
 PT binding by enterococci -
 XX
 PS Claim 26; Pages 119-122; 148pp; English.
 XX
 CC The present sequence is collagen binding Ace protein from Enterococcus
 CC faecalis, where Ace is adhesion of collagen from enterococcal bacteria.
 CC This protein is an extracellular matrix-binding protein, which can bind
 CC with collagens such as collagen type I and type IV and with lamins. Ace
 CC protein can be used to inhibit attachment of enterococci to collagen and
 CC to treat or prevent enterococcal infections, specifically as a vaccine.
 CC
 XX Sequence 458 AA:
 SQ
 Query Match 15.9%; Score 178; DB 21; Length 458;
 Best Local Similarity 28.3%; Pred. No. 5.9e-07;
 Matches 54; Conservative 33; Mismatches 88; Indels 16; Gaps 6;
 QY 18 TITSGNKSNTNVVHKSEAG---TSSVFYKTDMLPEDTHVWFLNNKSYSKDIT 74
 Db 121 tatatqrllegvntelqtglerdyfkyvgdlage-sqyrvflnvlnksdvredis 179
 QY 75 IKDQIGGQGLDSTLNTNVTGTHSNYVSGSALIDPEKAFPGSKRTVNTKNTIDVTIP 134
 Db 180 idrqqsggqqlnhesitfdvndketky--islaeteggygkldfv--tdndfnlrfy 234
 QY 135 OGYSNSFSINYSKTKIT--NEQKKEFVNNSQAWYOEHGKEEVNGKSPNHTVHNINANG 192
 Db 235 rdkarftsfivrystltaeqghatfensydlnglndatnektsgv-----knvf 289
 QY 193 IEGTVKGEILV 203
 Db 290 vegaasqngnv 300

RESULT 6
 AAB37668
 ID AAB37668 standard; protein; 146 AA.
 XX
 AC AAB37668;
 XX
 DT 28-FEB-2001 (first entry)
 XX
 DE Collagen binding Ace protein collagen binding domain.
 XX
 KM Collagen binding Ace protein; antibacterial; immunostimulant; vaccine;
 KW extracellular matrix-binding protein; lamin binding; bacterial infection.
 XX
 OS Enterococcus faecalis.
 XX
 PN WO200068242-A1.
 PD 16-NOV-2000.
 XX
 PF 10-MAY-2000; 2000WO-US12590.
 XX
 PR 10-MAY-1999; 99US-0133334.
 XX
 PA (TEXA) UNIV TEXAS A & M SYSTEM.
 PA (TEXA) UNIV TEXAS MEDICAL SCHOOL.
 XX

PI Rich RL, Kriekemeyer B, Owens RT, Hook M, Murray BE;
 PI Mallapareddy SR, Qin X, Weinstein GM, Singh KV, Duh R;
 XX WPI; 2000-687639/67.
 DR
 XX
 XX New collagen-binding protein from *Enterococcus*, useful e.g. in
 PT protective vaccines, for diagnosis and treatment of *Enterococcal*
 PT infections and for screening for compounds that inhibit collagen
 PT binding by *enterococci* -
 XX
 XX Disclosure; Fig 1; 148pp; English.
 PS
 XX The present invention relates to collagen binding Ace protein from
 CC *Enterococcus faecalis*, where Ace is adhesion of collagen from
 CC enterococcal bacteria (see AAC3828 and AAB3667). Ace protein is an
 CC extracellular matrix-binding protein, which can bind with collagens such
 CC as collagen type I and type IV and with lamins. Ace protein can be used
 CC to inhibit attachment of *enterococci* to collagen and to treat or prevent
 CC enterococcal infections, specifically as a vaccine. The present sequence
 CC is the collagen binding domain of Ace protein.
 CC
 XX Sequence 146 AA:
 SQ
 Query Match 14.8%; Score 165.5; DB 21; Length 146;
 Best Local Similarity 30.2%; Pred. No. 1.4e-06;
 Matches 45; Conservative 30; Mismatches 65; Indels 9; Gaps 5;
 QY . 41 FYYKTGDMLEPDTTHVWFLINNEKSYVSKDITIKDQIGGQQLDLSTLINVTGTHSN 100
 Db . 5 ffykvgdlage-sngyrvflnvnlnksdvtdedtsldrgsqgqlnkeftfdvndket 63
 QY 101 YVSGGSAITDFEKAFFGSKRTVDNTKNTIDVTIPQIGSYNSFSINVKTKIT--NEQKE 158
 Db . 64 ky---lslaefegqygkldfv---lndnflrfyrdkarktsfivrytstltaeqnqat 118
 QY 159 FVNSQAWYOEHGKEVNGKSFNHTVHNI 187
 Db 119 fensydingqlmqdatneknstq-vknv 146
 DE Fibronectin-binding MSCRAMM derivative pCF33.
 XX
 XX Fibronectin: pCF33; collagen binding protein; sepsis; infection;
 KW microbial surface component recognising adhesive matrix molecule;
 KM MSCRAMM; adhesin; vaccine; immunisation; diagnosis; therapy;
 KM mastitis.
 XX
 XX Staphylococcus aureus.
 OS
 XX
 XX Key Location/Qualifiers
 FH Peptide 1..12
 FT /note="vector pQE30-derived peptide"
 FT
 XX
 XX WO9743314-A2.
 PN
 XX 20-NOV-1997.
 PD
 XX 14-MAY-1997; 97WO-US08210.
 PF
 XX 16-MAY-1996; 96US-0017678.
 PR
 XX (UABR-) UAB RES FOUND.
 PA (TEXA) UNTV TEXAS A & M SYSTEM.
 PA
 XX

PI Hook M, House-Pompeo K, Patti JM, Sthanam N, Symersky J;
 XX WPI; 1998-008801/01.
 DR
 XX
 XX Antibody that interacts with collagen binding domain of
 PT *Staphylococcal* cna gene product - useful to prevent bacterial sepsis
 PT in animal infected with *Staphylococcus aureus*
 XX
 XX Disclosure; Page 91; 143pp; English.
 PS
 XX This protein comprises *Staphylococcus aureus* fibronectin-binding
 CC microbial surface component recognising adhesive matrix molecule
 CC (MSCRAMM) derivative pCF33, plus a vector-derived N-terminal
 CC peptide. The invention relates to claimed nucleic acid sequences
 CC (see AAT93436-38) encoding *S. aureus* collagen binding protein (CBP)
 CC epitopes M17, M31 and M55 (see AAW31552-54) that confer protection
 CC against *S. aureus* infection. CBP protein and antigenic epitopes
 CC are contemplated for use in the treatment of pathological
 CC infections, especially to prevent bacterial adhesion to collagen.
 CC The epitopes are also contemplated for use in the preparation of
 CC vaccines and as carrier proteins in vaccine formulations, as well
 CC as in the formulation of compositions for the prevention of *S.*
 CC *aureus* infection. pCF33 and pQD (see AAW31556) were used to raise
 CC anti-MSCRAMM polyclonal antibodies used in passive immunisation
 CC against bovine mastitis.
 CC
 XX Sequence 345 AA:
 SQ
 Query Match 11.2%; Score 125.5; DB 19; Length 345;
 Best Local Similarity 30.1%; Pred. No. 0.01; Mismatches 73; Indels 37; Gaps 10;
 Matches 56; Conservative 20;
 QY 1 MRGSHHHHNGS---DDKVAITTSGNKSTNV-----TVHKSSEAGTSVFFYKTKDM 48
 Db 1 mrgshhhhhngsmvaadapaagctdntqnlcnvctvgdsqctvphqgykloy---gfs 57
 QY 49 LPEDTTHVRWF-----LINNEKSYVS-KDITIKDQIGGQQLDLSTLINVTGTHSN 100
 Db 58 vpsavkvgdtkltvpkeInlgtvstakvypimaqgqvlangvidsdg---nvlyftld 114
 QY 101 YVSGGSAITDFEKA---FPGSKITVDNTKNTIDVTIPQIGSYNSFSINVKTKITNEQK 157
 Db 115 yvntk-----dvkactlmp-ayldpennvkktgnvclatqigstla---nktvldvdyeky 166
 QY 158 EFVNS 163
 Db 167 kfylns 172
 DE Clostridium botulinum type G toxin C fragment.
 XX
 XX Clostridium botulinum serotype G strain 113/30.
 OS
 XX Synthetic.
 OS
 XX Key Location/Qualifiers
 FH Peptide 1..21
 FT /note="N-terminal His tag"
 FT
 XX
 XX WO9808540-A1.
 PN
 XX 05-MAR-1998.
 PD

```

XX 28-AUG-1997; 97WO-US15394.
PF
XX
PR 28-AUG-1996; 96US-0704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Thalley BS, Williams JA;
XX
DR MPI: 1998-230234/20.
DR N-PSDB: AAV30596.
XX
PR Host cell containing recombinant expression vector encoding
XX Clostridium botulinum type B or E toxin - useful to treat humans
XX and other animals at risk of intoxication with clostridial toxin
XX
PS Example 49; Page 376-378; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment
CC of Clostridium botulinum (113/30 strain) type G neurotoxin, encoded
CC by a DNA sequence (see AAV30596) in plasmid pETN15b. This vector
CC can be used to express BotG soluble C fragment in Escherichia
CC coli host cells, with the recombinant C fragment being purified on
CC an affinity column. The invention relates to recombinant proteins
CC derived from C. botulinum toxins, especially type B and type E
CC toxins. Methods are provided which allow for the isolation of
CC soluble recombinant proteins free of significant endotoxin
CC contamination. Preferred hosts for production of recombinant
CC proteins are E. coli, insect cells and yeast cells. The
CC recombinant toxins are used as immunogens for the production of
CC vaccines and antitoxins that are useful in the treatment of humans
CC and animals at risk of intoxication with clostridial toxin.
XX
SQ Sequence 473 AA:

```

Query Match 10.3%; Score 116; DB 19; Length 473;
 Best Local Similarity 22.2%; Pred. No. 0.096; Mismatches 75; Indels 88; Gaps 13;
 Matches 56; Conservative 33;

```

QY 5 HHHHHGSDPKATVITSGKSTNVVHKSEAGTSVFYKGTGDMLEPETHYRWFPLINN 64
DB 6 hhhhhhsghl-----egfhmasma-----dlilqvf--nn 36
QY 65 EKSYSKDIITKDIQGGQDLDLSTLNINVTGTHSNYSGSARIDFEKAPFGSKTI 120
DB 37 ysnissnaallsyrgyllidsygaltnvgsdvlfnidngqfklhnse---nsm 92
QY 121 TVDNKNTIDVITPGGYGS-VNSFSIN-----KTIITNEQKEF-----VNSQAW 166
DB 93 fahgskfvv-----ydsmfafsfntwrtprkyunnidqlylqneyfliscikndsgw 145
QY 167 -YOEHGKE-----EVNGKS-----FNHTVNI--NANAGIEGT 196
DB 146 kvsignrliwrlidvnaakskeifeyskidnsdylnkwfsiltlndlganlyings 205
QY 197 VKGELKVLKQDK 208
DB 206 lksekilnldtr 217

```

RESULT 9
 AAY00241
 ID AAY00241 standard; Protein; 688 AA.
 AC AAY00241;
 XX
 XX 20-APR-1999 (first entry)
 XX Enterococcus faecalis antigenic polypeptide fragment EF124.
 DE Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
 XX detection; attenuation; antigenic.
 KW

```

XX OS Enterococcus faecalis.
XX
XX PN W09850554-A2.
XX
XX PD 12-NOV-1998.
XX
XX PF 04-MAY-1998; 98WO-US08959.
XX
XX PR 14-NOV-1997; 97US-0066009.
XX PR 06-MAY-1997; 97US-0044031.
XX PR 16-MAY-1997; 97US-0046655.
XX
XX PA (HDMA-) HDMA GENOME SCI INC.
XX
XX PI Bailey C, Choi GH, Hromocky J A, Kunsch CA;
XX
XX DR MPI: 1999-070095/06.
XX DR N-PSDB: AAX20231.
XX
XX PR New isolated Enterococcus faecalis polynucleotides - used to develop
XX products for the detection of Enterococcus and for use in vaccines
XX for prevention or attenuation of Enterococcus infection
XX
XX PS Claim 9; Page 230; 301pp; English.
XX
XX CC The present sequence represents an antigenic polypeptide fragment
XX isolated from Enterococcus faecalis. The present invention describes
XX genes, proteins and antigenic polypeptides isolated from E. faecalis.
XX The proteins can be used in vaccines for preventing or attenuating an
XX infection caused by a member of the Enterococcus genus in an animal.
XX They can also be used for detecting Enterococcus antibodies in a sample.
XX The nucleotide sequences can be used for detecting Enterococcus nucleic
XX acids. Products from the present invention can also be used for
XX screening compounds to identify agonists and antagonists of E. faecalis
XX protein activity.
XX
SQ Sequence 688 AA:

```

Query Match 10.2%; Score 114.5; DB 20; Length 688;
 Best Local Similarity 25.5%; Pred. No. 0.22; Mismatches 93; Indels 29; Gaps 9;
 Matches 54; Conservative 36;

```

QY 12 SDDKVAITISGNKSTNVVHKSEAGTSVFYKGTGDMLEPETHYRWFPLINNEKSYVS 70
DB 172 sydnasylngssrdvtcgvskibhgsesv--kgeylnkddpbdhyvwmlyngags-vl 228
QY 71 KQITTKDQIQGGQDLDLSTLNINVTGTHSNYSGSARIDFEKAPFGSKTIYDNKTN-TI 129
DB 229 ddvvltdlpspnqvlapeslvi--ygtl-----vledgtlirpdksvllleegkycl 277
QY 130 DVTIPGGYSYN-----SFSINXKTIITNEQ--QKEFVNNSQAWYOEHGKEEVNG 177
DB 278 evrltneitgqgkivkmaheapyymeysrlvtsaagsdclvsn-qvasltngseeving 336
QY 178 KSFNHTVININANAGIEGTGKGLKVLKQDK 209
DB 337 ddngdvvdvdhsgbhatgkglqkktamd 368

```

RESULT 10
 AAY00238
 ID AAY00238 standard; Protein; 2032 AA.
 AC AAY00238;
 XX
 XX 20-APR-1999 (first entry)
 XX Enterococcus faecalis protein EF123.
 DE Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
 XX detection; attenuation; antigenic.
 KW

```
XX OS Enterococcus faecalis.
XX PN W09850554-A2.
XX PD 12-NOV-1998.
XX PE 04-MAY-1998; 98MO-US08959.
XX PR 14-NOV-1997; 97US-0066009.
XX PR 06-MAY-1997; 97US-0044031.
XX PR 16-MAY-1997; 97US-0046655.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Bailey C, Choi GH, Hromocky J A, Kunsch CA;
XX WPI; 1999-070095/06.
XX DR N-PSDB; AAX20228.
XX PT New isolated Enterococcus faecalis polynucleotides - used to develop
XX PT products for the detection of Enterococcus and for use in vaccines
XX PT for prevention or attenuation of Enterococcus infection
XX PS Claim 9; Page 224-225; 301pp; English.
XX CC The present sequence represents a protein isolated from
XX CC Enterococcus faecalis. The present invention describes genes, proteins
XX CC and antigenic polypeptides isolated from E. faecalis. The proteins can
XX CC be used in vaccines for preventing or attenuating an infection caused
XX CC by a member of the Enterococcus genus in an animal. They can also be
XX CC used for detecting Enterococcus antibodies in a sample. The nucleotide
XX CC sequences can be used for detecting Enterococcus nucleic acids.
XX CC Products from the present invention can also be used for screening
XX CC compounds to identify agonists and antagonists of E. faecalis protein
XX CC activity.
XX SQ Sequence 2032 AA;

Query Match 10.2%; Score 114.5; DB 20; Length 2032;
Best Local Similarity 25.5%; Pred. No. 0.91;
Matches 54; Conservative 36; Mismatches 93; Indels 29; Gaps 9;

QY 12 SDDKVAATITSGNKSNTVTYHKS-EAGTSSVFYRTGDMLPEDTHVRFMLINNEKSYVS 70
DB 908 sydntlaasytngssrdvtygkvsidhgsesv--kkgyeyhkddpdlhyvwhmlngags-vl 964
QY 71 KDITIKQIOGGQQLDSTLININTGTHSNYSGOSATDFEKAFFPSKTIYDNTKN-TI 129
DB 965 ddvvvltclpnpqyldpeslvi--ygtln-----vtedqtltpdksvllleegkdytl 1013
QY 130 DVITIPQGYGSYN-----SFSINKTKITNEQ--QKEFVNNSQAWYQEHGKEEYNG 177
DB 1014 evtldnetgqkivvkmahleapymeyrslvtssaagstlvsn-qvsltngngsevvhg 1072
QY 178 KSFNHTVHNINANAGIECTVKGELKVLKQDKD 209
DB 1073 ddngdvvvldhsgshgahgtgkylqkkktamd 1104

RESULT 11
AAI00240
ID AAY00240 standard; Protein; 2032 AA.
XX
XX AAY00240;
XX
XX 20-APR-1999 (first entry)
XX DE Enterococcus faecalis protein EPI24.
XX DE Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
XX KW detection; attenuation; antigenic.
```

```
XX OS Enterococcus faecalis.
XX PN W09850554-A2.
XX PD 12-NOV-1998.
XX PE 04-MAY-1998; 98MO-US08959.
XX PR 14-NOV-1997; 97US-0066009.
XX PR 06-MAY-1997; 97US-0044031.
XX PR 16-MAY-1997; 97US-0046655.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Bailey C, Choi GH, Hromocky J A, Kunsch CA;
XX WPI; 1999-070095/06.
XX DR N-PSDB; AAX20230.
XX PT New isolated Enterococcus faecalis polynucleotides - used to develop
XX PT products for the detection of Enterococcus and for use in vaccines
XX PT for prevention or attenuation of Enterococcus infection
XX PS Claim 9; Page 228-229; 301pp; English.
XX CC The present sequence represents a protein isolated from
XX CC Enterococcus faecalis. The present invention describes genes, proteins
XX CC and antigenic polypeptides isolated from E. faecalis. The proteins can
XX CC be used in vaccines for preventing or attenuating an infection caused
XX CC by a member of the Enterococcus genus in an animal. They can also be
XX CC used for detecting Enterococcus antibodies in a sample. The nucleotide
XX CC sequences can be used for detecting Enterococcus nucleic acids.
XX CC Products from the present invention can also be used for screening
XX CC compounds to identify agonists and antagonists of E. faecalis protein
XX CC activity.
XX SQ Sequence 2032 AA;

Query Match 10.2%; Score 114.5; DB 20; Length 2032;
Best Local Similarity 25.5%; Pred. No. 0.91;
Matches 54; Conservative 36; Mismatches 93; Indels 29; Gaps 9;

QY 12 SDDKVAATITSGNKSNTVTYHKS-EAGTSSVFYRTGDMLPEDTHVRFMLINNEKSYVS 70
DB 908 sydntlaasytngssrdvtygkvsidhgsesv--kkgyeyhkddpdlhyvwhmlngags-vl 964
QY 71 KDITIKQIOGGQQLDSTLININTGTHSNYSGOSATDFEKAFFPSKTIYDNTKN-TI 129
DB 965 ddvvvltclpnpqyldpeslvi--ygtln-----vtedqtltpdksvllleegkdytl 1013
QY 130 DVITIPQGYGSYN-----SFSINKTKITNEQ--QKEFVNNSQAWYQEHGKEEYNG 177
DB 1014 evtldnetgqkivvkmahleapymeyrslvtssaagstlvsn-qvsltngngsevvhg 1072
QY 178 KSFNHTVHNINANAGIECTVKGELKVLKQDKD 209
DB 1073 ddngdvvvldhsgshgahgtgkylqkkktamd 1104

RESULT 12
AAI00242
ID AAY00242 standard; Protein; 2032 AA.
XX
XX AAY00242;
XX
XX 20-APR-1999 (first entry)
XX DE Enterococcus faecalis protein EPI25.
XX DE Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
XX KW detection; attenuation; antigenic.
```



```

Db      1 mrgshhhhhgscqalgqdmvspcatnsssfsspsagrhvrsynhlgd---vrvrk 57
QY      61 NINKEYSYVKDITKDQIOGSGQ-----LDLSTLNINVTG---HSWYSGQSAITD 110
Db      58 lsfktyflk--lekngkvsqtkkencpyslleltsvelgvvavkalnsny---lamnk 112
QY      111 FEAFPGSKITVD-NRKNTIDVTIPQGYGSYNSFSINYKTKRTNEQKKEPVNNSQAWYOE 169
Db      113 kqlllysketndcklkerie---engyntyaafn-----wqn 147
QY      170 HGKE---EVNGK 178
Db      148 ngrgmyvalngk 159

RESULT 14
AAB10309
ID      AAB10309 standard; Protein; 184 AA.
XX
XX      AAB10309;
AC
XX      20-NOV-2000 (first entry)
DT
XX
XX      Human KGF-2 construct pQE60-Cys37 protein SEQ ID NO: 30.
DE
XX
XX      Human; keratinocyte growth factor; KGF-2; antiulcer; antidiabetic;
KW      antiinflammatory; cytoprotective; dermatological; gastrointestinal;
KW      hepatic; respiratory; renal; cerebroprotective; mucositis; treatment;
KW      epithelial cell proliferation; inflammatory bowel disease; lung damage;
KW      liver disorder; diabetes; oral injury; gastrointestinal injury;
KW      gut toxicity; gastric; duodenal; epidermolysis bullosa; skin graft;
KW      skin disorder; renal failure; brain injury; intestinal fibrosis;
KW      proctitis; female reproductive tract disorder; pulmonary fibrosis;
KW      pneumonitis; pleural retraction; hemopoietic syndrome; myelotoxicity.
XX
XX      Homo sapiens.
OS
XX
XX      US6077692-A.
PN
XX      20-JUN-2000.
PD
XX
XX      13-FEB-1998; 98US-0023082.
PF
XX
XX      13-AUG-1996; 96US-0023852.
PR      28-FEB-1997; 97US-0039045.
PR      13-AUG-1997; 97US-0055561.
PR      05-JUN-1995; 95US-0461195.
PR      14-FEB-1995; 95WO-US01790.
PR      23-MAY-1997; 97US-0862432.
PR      13-AUG-1997; 97US-0910875.
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX      Mendrick D, Duan DR, Ni J, Jimenez P, Coleman TA, Gruber JR;
PI      Dillon PJ, Gentz RL, Ruben SM, Zhang J, Moore PA, Rampy MA;
XX
XX      WPI: 2000-441307/38.
DR      N-PSDB; AAA71215.
DR
XX
XX      Novel keratinocyte growth factor useful for promoting and accelerating
PT      wound healing, comprising at least 10 contiguous amino acids from a
PT      specific amino acid sequence -
XX
XX      Example 7; Fig 15; 190pp; English.
XX
XX      This invention describes a novel human keratinocyte growth factor, KGF-2
CC      (I), which has antiulcer, antidiabetic, antiinflammatory, cytoprotective,
CC      dermatological, gastrointestinal, hepatic, respiratory, renal and
CC      cerebroprotective activity. (I) is useful for stimulating epithelial cell
CC      proliferation in patients suffering from wound, mucositis, ulcer such as
CC      venous stasis ulcer, diabetic ulcer and cubitus ulcer. (I) is also useful
CC      for treating inflammatory bowel disease, liver disorder, lung damage,
CC      diabetes, oral injury, gastrointestinal injury, gut toxicity, gastric

```

```

CC      ulcer, duodenal ulcer, epidermolysis bullosa, skin graft, skin disorder,
CC      renal failure, brain injury, breast tissue injury, urothelial damage,
CC      female reproductive tract disorder, intestinal fibrosis, proctitis,
CC      pulmonary fibrosis, pneumonitis, pleural retraction, hemopoietic syndrome
CC      and myelotoxicity. (I) is also useful for increasing the adherence of
CC      skin grafts to wound beds and to stimulate re-epithelialization from the
CC      wound bed, to produce changes in hepatocyte proliferation, to reduce the
CC      side effects of gut toxicity, to regenerate skin in full and partial
CC      thickness skin defects, and to prevent and heal damage to lungs. KGF-2
CC      shows enhanced activity, increased stability, higher yield and better
CC      solubility. This sequence represents the human KGF-2 protein construct
CC      pQE60-Cys37 described in the method of the invention.
XX
XX      Sequence 184 AA;
SQ

Query Match      10.1%; Score 113.5; DB 21; Length 184;
Best Local Similarity 23.4%; Pred. No. 0.045;
Matches 45; Conservative 35; Mismatches 65; Indels 47; Gaps 9;

QY      1 MRSHHHHHSDKVAATITSGNKSITNVYHKSEAGTSVYYKTDLPDPTTHVWFL 60
Db      1 mrgshhhhhgscqalgqdmvspcatnsssfsspsagrhvrsynhlgd---vrvrk 57
QY      61 NINKEYSYVKDITKDQIOGSGQ-----LDLSTLNINVTG---HSWYSGQSAITD 110
Db      58 lsfktyflk--lekngkvsqtkkencpyslleltsvelgvvavkalnsny---lamnk 112
QY      111 FEAFPGSKITVD-NRKNTIDVTIPQGYGSYNSFSINYKTKRTNEQKKEPVNNSQAWYOE 169
Db      113 kqlllysketndcklkerie---engyntyaafn-----wqn 147
QY      170 HGKE---EVNGK 178
Db      148 ngrgmyvalngk 159

RESULT 15
AAB61673
ID      AAB61673 standard; Protein; 184 AA.
XX
XX      AAB61673;
AC
XX      10-APR-2001 (first entry)
DT
XX
XX      pQE60-Cys37 construct protein.
DE
XX
XX      Keratinocyte growth factor; KGF-2; epithelial cell proliferation; wound;
KW      mucositis; ulcer; inflammatory bowel disease; liver disorder;
KW      lung damage; diabetes; oral injury; gastrointestinal injury;
KW      epidermolysis bullosa; renal failure; brain injury; proctitis;
KW      pulmonary fibrosis; haemopoietic syndrome; ovary injury; infertility;
KW      liver fibrosis.
XX
XX      Unidentified.
OS
XX
XX      WO200102433-A1.
PN
XX
XX      11-JAN-2001.
PD
XX
XX      03-JUL-2000; 2000WO-US18328.
PF
XX
XX      02-JUL-1999; 99US-0142343.
PR      14-JUL-1999; 99US-0143648.
PR      15-JUL-1999; 99US-0144024.
PR      12-AUG-1999; 99US-0148628.
PR      19-AUG-1999; 99US-0149935.
PR      03-NOV-1999; 99US-0163375.
PR      22-DEC-1999; 99US-0171677.
PR      19-APR-2000; 2000US-0198322.
PR      19-MAY-2000; 2000US-0205417.
PR      30-JUN-2000; 2000US-9911224.
XX

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2002, 23:13:57 ; Search time 65.03 Seconds
(without alignments)
73.015 Million cell updates/sec

Title: US-09-813-820-4
Sequence: 1 MRGSHHHHHSDSKVAIT.....GIEGVKGLAVLKDKDKTK 211
Perfect score: 1122

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
4: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
5: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
6: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1122	100.0	211	4	US-08-856-253-4
2	1054.5	94.0	512	4	US-08-856-253-6
3	1048	93.4	1183	2	US-08-447-031A-2
4	830	74.0	159	4	US-08-856-253-2
5	125.5	11.2	345	4	US-08-856-253-7
6	113.5	10.1	184	3	US-09-023-082A-30
7	110.5	9.8	1338	4	US-08-728-470-9
8	110.5	9.8	1338	4	US-08-719-641-9
9	110.5	9.8	1599	2	US-08-617-697-9
10	104.5	9.3	1529	2	US-08-728-470-10
11	104.5	9.3	1529	4	US-08-719-641-10
12	104.5	9.3	1600	2	US-08-617-697-10
13	103.5	9.2	139	4	US-08-856-253-8
14	100	8.9	1536	1	US-08-038-682-2
15	100	8.9	1536	1	US-08-302-832-2
16	100	8.9	1536	2	US-08-530-198-2
17	100	8.9	1536	2	US-08-469-880-2
18	100	8.9	1536	2	US-08-728-470-2
19	100	8.9	1536	2	US-08-617-697-2
20	100	8.9	1536	4	US-08-719-641-2
21	99.5	8.8	589	4	US-09-377-155-19
22	98.5	8.8	351	2	US-08-701-191A-3
23	95	8.5	666	2	US-08-737-716-14
24	91	8.1	599	1	US-08-172-331B-4
25	90	8.0	76	3	US-08-956-307B-17
26	90	8.0	76	3	US-08-956-307B-18
27	90	8.0	122	3	US-08-956-307B-19

28	90	8.0	2123	4	US-08-968-685A-10	Sequence 10, Appl
29	89.5	8.0	1702	4	US-08-296-791-5	Sequence 5, Appl
30	89.5	8.0	1702	5	PCT-US85-10661A-5	Sequence 5, Appl
31	89	7.9	599	2	US-09-032-315-6	Sequence 6, Appl
32	89	7.9	599	2	US-08-993-318A-6	Sequence 6, Appl
33	89	7.9	599	4	US-09-399-886-6	Sequence 6, Appl
34	89	7.9	599	4	US-09-396-260-6	Sequence 6, Appl
35	89	7.9	599	4	US-09-576-281-6	Sequence 6, Appl
36	89	7.9	631	1	US-08-605-541B-12	Sequence 12, Appl
37	87	7.8	254	4	US-09-004-731-89	Sequence 92, Appl
38	87	7.8	254	4	US-09-004-731-92	Sequence 92, Appl
39	87	7.8	254	4	US-08-749-699-89	Sequence 89, Appl
40	87	7.8	254	4	US-08-749-699-92	Sequence 89, Appl
41	86.5	7.7	2353	4	US-09-377-155-33	Sequence 33, Appl
42	86.5	7.7	2353	4	US-08-913-942-4	Sequence 4, Appl
43	86	7.7	216	2	US-08-737-716-4	Sequence 4, Appl
44	86	7.7	310	3	US-09-136-628-2	Sequence 2, Appl
45	86	7.7	437	2	US-08-737-716-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-856-253-4
Sequence 4, Application US/08856253
Patent No. 6288214
GENERAL INFORMATION:
APPLICANT: Hook, Magnus
APPLICANT: Patti, Joseph M.
APPLICANT: House-Pompeo, Karen
APPLICANT: Stahanam, Narayana
APPLICANT: Symersky, Jindrich
TITLE OF INVENTION: COLLAGEN BINDING PROTEIN COMPOSITIONS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: U.S.
ZIP: 77210
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856,253
CLASSIFICATION: 514
FILING DATE: Concurrently Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/017,678
FILING DATE: 16-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: TAMK:193
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 211 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-856-253-4
Query Match 100.0% Score 1122; DB 4; Length 211;
Best Local Similarity 100.0%; Pred. No. 2.1e-94;

	Matches	211; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
QY	1	MRGSHHHHHSDDKVAITTSGNKSTNTVYHKSEAGTSSVFYTKGDMIPEDTTHYRWF	60			
Db	1	MRGSHHHHHSDDKVAITTSGNKSTNTVYHKSAGTSSVFYTKGDMIPEDTTHYRWF	60			
QY	61	NINNEKSYVKDITFKDDIOGGQQLDELTLNLTNVGTGSHSNYSQGSAITDPFKAAPGSKI	120			
Db	61	NINNEKSYVKDITFKDDIOGGQQLDELTLNLTNVGTGSHSNYSQGSALTDPFKAAPGSKI	120			
QY	121	TVDNTRKNTIDVTIPGGYSYNFSFNYYTKKITNEOQKEFVNNSQAMYOEHKGEEYNGKF	180			
Db	121	TVDNTRKNTIDVTIPGGYSYNFSFNYYTKKITNEOQKEFVNNSQAMYOEHKGEEYNGKSF	180			
QY	181	NHTVHNANANGIECTVGELKVLKODDDTK	211			
Db	181	NHTVHNANANGIECTVGELKVLKODDDTK	211			

RESULT 2
 US-08-856-253-6
 Sequence 6, Application US/08856253
 Patent No. 6288214
 GENERAL INFORMATION:
 APPLICANT: Hook, Magnus
 APPLICANT: Patil, Joseph M.
 APPLICANT: House-Pompeo, Karen
 APPLICANT: Shanam, Narayana
 APPLICANT: Symersky, Jindrich
 TITLE OF INVENTION: COLLAGEN BINDING PROTEIN COMPOSITIONS
 TITLE OF INVENTION: AND METHODS OF USE
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Arnold, White & Durkee
 STREET: P.O. Box 4433
 CITY: Houston
 STATE: Texas
 COUNTRY: U.S.
 ZIP: 77210
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/856,253
 FILING DATE: Concurrently Herewith
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/017,678
 FILING DATE: 16-MAY-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Kitchell, Barbara S.
 REGISTRATION NUMBER: 33,928
 REFERENCE/DOCKET NUMBER: TAMK:193
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (512) 418-3000
 TELEFAX: (512) 474-7577
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 512 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 US-08-856-253-6

```

Query Match          94.0%;  Score 1054.5;  DB 4;  Length 512;
Best Local Similarity 64.7%;  Pred. No. 9,6e-88;
Matches 211;  Conservative 0;  Indels 115;  Gaps 1
OY      1 MRGSHNNHNG----- 11
        |||||

```

Db	1	MRGSHHHHHGCSADISTNVNTDLTVSPSKIEDGKTTVYKMFEDDKNGKIQNDMDIKVAM	60
QY	12	-----	11
Db	61	PTSGTVKIEGYSKTVPLTVKGEVGAATIPDCAITTFNDKVEKLSVGSFAEFVQGRN	120
QY	12	-----SDDKVATITSGNKSNTNVVHKSEAGTSSVFYKTKGDM,PEDTTHVRMFLINNE	65
Db	121	LTQNTSDDKVATITSGNKSNTNVVHKSEAGTSSVFYKTKGDM,PEDTTHVRMFLINNE	180
QY	66	KSYYSKDTTIDMOIQGGQOIDLSTLNTNVGTHSNYSSGSAITTFDEKAPGSKITVNDT	125
Db	181	KSYYSKDTTIDMOIQGGQOIDLSTLNTNVGTHSNYSSGSAITTFDEKAPGSKITVNDT	240
QY	126	KNTIDVTIPQGYGSNSFSINYKRTITNEOQKEFVNNSSQAMOEHGKEEYNGKSFNHTVH	185
Db	241	KNTIDVTIPQGYGSNSFSINYKRTITNEOQKEFVNNSSQAMOEHGKEEYNGKSFNHTVH	300
QY	186	NINANAGIEGTVKGELVYVKMDKRTK	211
Db	301	NINANAGIEGTVKGELVYVKMDKRTK	326

RESULT 3
 US-08-447-031A-2
 : Sequence 2, Application US/08447031A
 : Patent No. 5851794
 :
 : GENERAL INFORMATION:
 : APPLICANT: GUSS, Bengt
 : APPLICANT: HOOK, Magnus
 : APPLICANT: JONSSON, Hans
 : APPLICANT: LINDBERG, Martin
 : APPLICANT: PATRI, Joseph
 : APPLICANT: SIGMÄS, Christer
 : APPLICANT: SWITALSKI, Lech
 : TITLE OF INVENTION: A COLLAGEN BINDING PROTEIN AS WELL AS
 : TITLE OF INVENTION: ITS PREPARATION
 : NUMBER OF SEQUENCES: 8
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Burns, Doane, Swecker & Mathis
 : STREET: P.O. Box 1404
 : CITY: Alexandria
 : STATE: Virginia
 : COUNTRY: United States
 : ZIP: 22313-1404
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 :
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/447,031A
 : FILING DATE: 22-MAY-1995
 :
 : CLASSIFICATION: 435
 :
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/861,804
 : FILING DATE: 21-AUG-1992
 :
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: WO PCT/SE91/00707
 : FILING DATE: 22-OCT-1991
 :
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: SE 9003374-7
 : FILING DATE: 22-OCT-1990
 : ATTORNEY/AGENT INFORMATION:
 : NAME: McGowan, Malcolm K.
 : REGISTRATION NUMBER: 39,300
 : REFERENCE/DOCKET NUMBER: 012889-006
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (703) 836-6620
 : TELEFAX: (703) 836-2021
 : INFORMATION FOR SEQ ID NO: 2:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1183 amino acids


```

Db 1 MRGSHHHHGSVNAADAPAGTDTITNLTNTVIGIDSGTIVVPHOAGYKLVN---GFS 57
QY 49 LPEDTTHVME-----LNINNEKSYVS-KDTIKDOLOGGQQLDLSTLINIVGTGHSN 100
Db 58 VPRSAVKGTFKTVKRELNGVNTAKVPTMAGDQVLANGVDSG---NVITTFD 114
QY 101 YVSGQSAITDEKA---FPGSKITVDNTKNTIDVTIPQGYGYSNFSINXKTKITNEQOK 157
Db 115 YVNTK----DVYKATLTMF-AYIDPENVKKTGNVTLATIGISGTTA---KTVLVYDEKKG 166
QY 158 EFPNNS 163
Db 167 KEYNLS 172

RESULT 6
Sequence 30, Application US/09023082A
Patent No. 6077692
GENERAL INFORMATION:
APPLICANT: RUBEN, STEVEN M.
APPLICANT: JIMENEZ, PABLO
APPLICANT: DUAN, D. ROXANNE
APPLICANT: RAPPY, MARK A.
APPLICANT: MENDRICK, DONNA
APPLICANT: ZHANG, JUN
APPLICANT: NI, JIAN
APPLICANT: MOORE, PAUL A.
APPLICANT: COLEMAN, TIMOTHY A.
APPLICANT: GRUBER, JOACHIM R.
APPLICANT: DILLON, PATRICK J.
APPLICANT: GENTZ, REINER L.
TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVE, NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,082A
FILING DATE: 13-FEB-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/01790
FILING DATE: 14-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/461,195
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/023,852
FILING DATE: 13-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/039,045
FILING DATE: 28-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/862,432
FILING DATE: 23-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/910,875
FILING DATE: 13-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/055,561
FILING DATE: 13-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.

```

```

REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0360008/EKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 184 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-023-082A-30

Query Match
10.1%; Score 113.5; DB 3; Length 184;
Best Local Similarity 23.4%; Pred. No. 0.0046;
Matches 45; Conservative 35; Mismatches 65; Indels 47; Gaps 9;

QY 1 MRGSHHHHSDDKVATYTGSKNSTVTVHKSEAGTSSVFYKGTGDMLPEDTTHVMEFL 60
Db 1 MRGSHHHHSGQALGQDWVSPDATNSSSSPSSPSAGRHVRSYNHLOGD---VRMRK 57
QY 61 NINNEKSYVSKDTITDOLOGGQ-----LDLSTLINIVGT---HSNYVSGQSAITD 110
Db 58 LFSFTYFLK--LEKNGKVSQTKREKCPYSILEITVEIGVAVKAINSYV---LMMK 112
QY 111 FEKAFPGSKITVD-NTRKNIDVTIPQGYGYSNFSINXKTKITNEQOKKEFVNNSQAMYO 169
Db 113 KGKLYGSKERNDCIKERIE---ENGINTYASFN-----MCH 147
QY 170 HGKE---EVNGK 178
Db 148 NGRQMYVALNGK 159

RESULT 7
US-08-728-470-9
Sequence 9, Application US/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651

```

```
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1338 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-9

Query Match          9.8%; Score 110.5; DB 2; Length 1338;
Best Local Similarity 25.5%; Pred. No. 0.14;
Matches 61; Conservative 28; Mismatches 89; Indels 61; Gaps 10;

QY 12 SDDKVAITISGNKST-----NVTYHKSEA-----GTSVFEYKTDMLPEDTHVWFNLN 61
DB 694 SDTYRTITIKGNISNKSGLNIIIDKSDAEIQGNIS---QKEGN-----LT 738
QY 62 INNEKSVSKDITIKDQIQGGQQLDSTLNINVT-GTHSNYSGSALITDEKAFPGSKI 120
DB 739 ISSDKVNITNQITIKAGVEGGRSDSEAEANLTIQTKEKLKLAGDLNISGFNKA-----EI 794
QY 121 TVDNKNTIDVTIPQGYGYSY-----NSFSINYKTKITNEQKEFY 160
DB 795 T---AKNGSDLTIGNASGNADAKKVTFDKVKDSKISTDGHNVTLNSEVKTSGSSNAGN 851
QY 161 NNSQAMQOEHGKEEVNKSFNHTVHNINANAG----IEGT----VKSELKVLKQDKDTK 211
DB 852 DNSTGLTISAKDVTYNNNVNNTSHKTTINISAAAGVTTKEGTTINATTGSVEYTAQNGTIK 910

RESULT 8
US-08-719-641-9
Sequence 9, Application US/08719641
Patent No. 6218141
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Maltare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
```

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1338 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-9

Query Match          9.8%; Score 110.5; DB 4; Length 1338;
Best Local Similarity 25.5%; Pred. No. 0.14;
Matches 61; Conservative 28; Mismatches 89; Indels 61; Gaps 10;

QY 12 SDDKVAITISGNKST-----NVTYHKSEA-----GTSVFEYKTDMLPEDTHVWFNLN 61
DB 694 SDTYRTITIKGNISNKSGLNIIIDKSDAEIQGNIS---QKEGN-----LT 738
QY 62 INNEKSVSKDITIKDQIQGGQQLDSTLNINVT-GTHSNYSGSALITDEKAFPGSKI 120
DB 739 ISSDKVNITNQITIKAGVEGGRSDSEAEANLTIQTKEKLKLAGDLNISGFNKA-----EI 794
QY 121 TVDNKNTIDVTIPQGYGYSY-----NSFSINYKTKITNEQKEFY 160
DB 795 T---AKNGSDLTIGNASGNADAKKVTFDKVKDSKISTDGHNVTLNSEVKTSGSSNAGN 851
QY 161 NNSQAMQOEHGKEEVNKSFNHTVHNINANAG----IEGT----VKSELKVLKQDKDTK 211
DB 852 DNSTGLTISAKDVTYNNNVNNTSHKTTINISAAAGVTTKEGTTINATTGSVEYTAQNGTIK 910
```

```
RESULT 9
US-08-617-697-9
Sequence 9, Application US/08617697
Patent No. 5977336
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Maltare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
```

SEQUENCE CHARACTERISTICS:
LENGTH: 1599 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-9

Query Match 9.8%; Score 110.5; DB 2; Length 1599;
Best Local Similarity 25.5%; Pred. No. 0.18;
Matches 61; Conservative 28; Mismatches 89; Indels 61; Gaps 10;

QY 12 SDDKAVTITSGKST-----NTVHKSEA-----GTSVFYKTKGMLPBDTTHRWFL 61
DB 954 SPTTYRTIHKGNISNKGSLDIIDKKDAEIQIGNIS---QKEGN-----LT 998
QY 62 INNERSYVSKDITKIDQIGGOQLDLSTLINVT--GTHSNYSGSALIDFEKAFPGSK 120
DB 999 ISSDKVNIITNQTITKAGVGGSDSEAEANALITQTKELKLAGDLNISGFNKA-----E 1054
QY 121 TVDNTKNTIDVTIPOGYSY-----NSFSINYKTKITNEQOKEFY 160
DB 1055 T---AKNGSDLTIGNASGGMADAKKVTDPKVKDSKISTDGHVNTLNSVKTSNGSSNAG 1111
QY 161 NNSQAMVQEHGKEEVNGKSFNHTVHINANAG-----IEGT---VKGLKVLKQDKDTK 211
DB 1112 DNSTGLTISAKDVTYNNNTVSHRTINISAAAGVTTKEGTTINATGTSVEVTAQNGTIR 1170

RESULT 10

US-08-728-470-10

; Sequence 10, Application US/08728470

; Patent No. 5928651

; GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen J

; TITLE OF INVENTION: High Molecular Weight Surface Proteins

; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Shoemaker and Mattare, Ltd.

; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

; STREET: Bldg. 1

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202-0286

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/728,470

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/302,832

; FILING DATE: 16-MAR-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US PCT/US93/02166

; FILING DATE: 16-MAR-1993

; APPLICATION NUMBER: GB 9205704.1

; FILING DATE: 16-MAR-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Berkstreser, Jerry W

; REGISTRATION NUMBER: 22,651

; REFERENCE/DOCKET NUMBER: 1038-633

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 415-0810

; TELEFAX: (703) 415-0813

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-10

Query Match 9.3%; Score 104.5; DB 2; Length 1529;
Best Local Similarity 22.9%; Pred. No. 0.58;
Matches 55; Conservative 32; Mismatches 90; Indels 63; Gaps 9;

QY 16 VALITSGKST-----NTVHKSEAGTSSV-----FYKTDMLPBDTTHRWFL 60
DB 881 ITTNAGTQRTIINGNITKEKGLNLIKADAEIQIGNISQKEN-----L 928
QY 61 INNERSYVSKDITKIDQIGGOQLDLSTLINVT--GTHSNYSGSALIDFEKAFPGSK 119
DB 929 TISSDKVNIITNQTITKAGVGGSDSEAEANALITQTKELKLAGDLNISGFNKA-----E 984
QY 120 TVDNTKNTIDVTIPOGYSY-----NSFSINYKTKITNEQOKEFY 159
DB 985 IT---AKNGSDLTIGNASGGMADAKKVTDPKVKDSKISTDGHVNTLNSVKTSNGSSNAG 1041
QY 160 VNSQAMVQEHGKEEVNGKSFNHTVHINANAG-----IEGT---VKGLKVLKQDKDTK 211
DB 1042 DNSTGLTISAKDVTYNNNTVSHRTINISAAAGVTTKEGTTINATGTSVEVTAQNGTIR 1101

RESULT 11

US-08-719-641-10

; Sequence 10, Application US/08719641

; Patent No. 6218141

; GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen J

; TITLE OF INVENTION: High Molecular Weight Surface Proteins

; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Shoemaker and Mattare, Ltd.

; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

; STREET: Bldg. 1

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202-0286

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/719,641

; FILING DATE:

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/302,832

; FILING DATE: 16-SEP-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US PCT/US93/02166

; FILING DATE: 16-MAR-1993

; APPLICATION NUMBER: GB 9205704.1

; FILING DATE: 16-MAR-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Berkstreser, Jerry W

; REGISTRATION NUMBER: 22,651

; REFERENCE/DOCKET NUMBER: 1038-625

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 415-0810

; TELEFAX: (703) 415-0813

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1529 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-10

Query Match 9.3%; Score 104.5; DB 4; Length 1529;
Best Local Similarity 22.9%; Pred. No. 0.58;
Matches 55; Conservative 32; Mismatches 90; Indels 63; Gaps 9;

QY 16 VAITSGNKT-----NVTVKSEAGTSSV-----FYKKGMDLPEDTHVWFL 60
DB 881 ITTNASGTOQTIIINGNTNEKGLNKNIKADAEIQIGNSQKEGN-----L 928
QY 61 NINNEKSYVSKDITIKDQIOGGOOLDLSTLINVT-GTHSNYSGQSAITDFEKAPGSK 119
DB 929 TISSDKVNITNOITIKAGVGRSDSEANMNLTIOTKELKLAGDLNISGFNKA-----E 984
QY 120 ITVDNKNITDVTIPOGYSY-----NSFSINYKRTINEOKEF 159
DB 985 IT---AKNSGDLITIGNASGGMADAKKVTDFKVKDSKISTDGHNVTLNSEVKTSSNGSSNAG 1041
QY 160 VNNSQAMYOEHGKEEYNGKSFNHTVHINANAG---IEGT---VKGELKVLKQDKDK 211
DB 1042 NDNSTGLTISAKDVTYNNVNTSHKTTINISAAAGVTTKEGTTINATGSEVYAQNQGTIK 1101

RESULT 12
US-08-617-697-10
Sequence 10, Application US/08617697
Patent No. 5977336
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1600 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-10

Query Match 9.3%; Score 104.5; DB 2; Length 1600;
Best Local Similarity 22.9%; Pred. No. 0.62;
Matches 55; Conservative 32; Mismatches 90; Indels 63; Gaps 9;

QY 16 VAITSGNKT-----NVTVKSEAGTSSV-----FYKKGMDLPEDTHVWFL 60
DB 951 ITTNASGTOQTIIINGNTNEKGLNKNIKADAEIQIGNSQKEGN-----L 998
QY 61 NINNEKSYVSKDITIKDQIOGGOOLDLSTLINVT-GTHSNYSGQSAITDFEKAPGSK 119
DB 999 TISSDKVNITNOITIKAGVGRSDSEANMNLTIOTKELKLAGDLNISGFNKA-----E 1054
QY 120 ITVDNKNITDVTIPOGYSY-----NSFSINYKRTINEOKEF 159
DB 1055 IT---AKNSGDLITIGNASGGMADAKKVTDFKVKDSKISTDGHNVTLNSEVKTSSNGSSNAG 1111
QY 160 VNNSQAMYOEHGKEEYNGKSFNHTVHINANAG---IEGT---VKGELKVLKQDKDK 211
DB 1112 NDNSTGLTISAKDVTYNNVNTSHKTTINISAAAGVTTKEGTTINATGSEVYAQNQGTIK 1171

RESULT 13
US-08-856-253-8
Sequence 8, Application US/08856253
Patent No. 6288214

GENERAL INFORMATION:
APPLICANT: Hook, Magnus
APPLICANT: Patti, Joseph M.
APPLICANT: House-Pompeo, Karen
APPLICANT: Stnam, Narayana
APPLICANT: Symersky, Jindrich
TITLE OF INVENTION: COLLAGEN BINDING PROTEIN COMPOSITIONS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: U.S.
ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856,253
FILING DATE: Concurrently Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/017,678
FILING DATE: 16-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: TAMK:193
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 139 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-856-253-8

Query Match 9.2%; Score 103.5; DB 4; Length 139;
Best Local Similarity 25.7%; Pred. No. 0.025;

```
Matches      46; Conservative    16; Mismatches     54; Indels       63; Gaps        9;

QY   1 MRGSHHHHSGDDKVAITTSGNKSTNVTVHKSEAGTSSVFYYKTGDMLPEDTHVRMFL 60
D8   1 MRGSHHHHHSGE---GGNSNGNOS-----FEEDTE----- 28
QY   61 NINNEKSYSVMDITIKDOIGQGQLDLSTLININVGTGSNTSYSGSQAITEPKAPPGSKI 120
Db   29 --EDKPKE-----OGGVNYDIIDF--DSYQHIGNKGKNOSFEDEEK----DKP 70
QY   121 TVDNTKATIDV---TIPOGYSYSNFISINYKYKITNEOOKREFNNASOAWOEHRGREEN 176
Db   71 KYEHGNIIDIIDFDPSVPRIHG-----FNKHREILED---TNNDKFSYORGCHNSVD 119

RESULT      14
US-08-038-682-2
Sequence 2, Application US/08038682
Patent No. 5549897
GENERAL INFORMATION:
APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GEME III, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHIIDS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Matlare, Ltd
STREET : 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
          STREET: Bldg. 1
CITY : Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MC-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/038,682
FILING DATE: 16-MAR-1993
CLASSIFICATION: 42A
ATTORNEY/AGENT INFORMATION:
NAME: BERKSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-293
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-038-682-2

Query Match           8.9%; Score 100; DB 1; Length 1536;
Best Local Similarity 25.2%, Pred. No. 1.5;
Matches 51; Conservative 33; Mismatches 72; Indels 46; Gaps 9.
```

```

Db   1057 DGSULTIGNT-NSADGT-----NAKKVFNQVKOSKISADGHKVTL-----HS 1098
Oy   172 KEEVNGKSFNHTVNINANAGI 193
      | : | | : | : | : | : |
Db   1099 KVETSG-SNNNTEDSDNNAGL 1119

RESULT 15
US-08-302-832-2
; Sequence 2, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0810
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-302-832-2

Query Match          8.9%; Score 100; DB 1; Length 1536;
Best Local Similarity 25.2%; Pred. No. 1.5;
Matches    51; Conservative 33; Mismatches    72; Indels    46; Gaps    9;

Oy   13 DDKVAITTSGKNSTNVTVHKSEAGTSSVFYKTDGM-----LPEDTHVRNFL 60
      | : | | : | | : | : | | | : | : |
Db   943 NSKNSTISTNSSSYRTI-----ISGNITKNGDLINTNEGSDTEMOIGGVSOKEGL 996

Oy   61 NNNEKSIVSKDITKDIOGGQQDLSTLNINNV-GTHSNVYSQSASATIDPEK---AF 115
      | : | | : | | | : | : | | | : | : |
Db   997 TTSSDKINIKQTITKACAVDGENSDSATNNANLTIKTKELKLITODLINSGFNKETIAK 1056

Oy   116 PGSKTVTNKTITDIPTIPOGYGSYNFSINYK---TKITNEOQKEFVNNSQAWYOEHG 171
      | : | : | : | : | : | : | : | : | : |
Db   1057 DGSULTIGNT-NSADGT-----NAKKVFNQVKDSKISADGHKVTL-----HS 1098

Oy   172 KEEVNGKSFNHTVNINANAGI 193
```

Thu Jan 31 10:59:13 2002

us-09-813-820-4.rail

Page 9

Db 1099 KVEISG-SNNTEDESSDNNAGL 1119

Search completed: January 30, 2002, 00:02:24
Job time: 2907 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2002, 00:02:27 ; Search time 49.08 Seconds

(without alignments)
157.626 Million cell updates/sec

Title: US-09-813-820-4

Perfect score: 1122
Sequence: 1 MRGSHHHHHGSDKVAFT.....GIEGVKGLKVLKQDKDTK 211

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1048	93.4	1183	1 CNA_STAU	Q53654 staphylococ
2	106	9.4	1076	1 HSER_CAVPO	P70106 cavia porce
3	98	8.7	1256	1 ATL_STAU	P52081 staphylococ
4	96.5	8.6	848	1 TOP1_BORBU	O51768 borrelia bu
5	96	8.6	704	1 MSN2_YEAST	P33748 saccharomyc
6	95	8.5	666	1 MUR2_ENTHR	P39046 enterococcu
7	95	8.5	2003	1 YDBA_ECOLI	P33666 escherichia
8	93.5	8.3	1296	1 BXG_CLOBO	P36666 clostridium
9	92.5	8.2	711	1 TOP1_MYCPN	P78032 mycoplasma
10	91.5	8.2	238	1 OPAH_NRIGO	O04884 neisseria g
11	91	8.1	463	1 GUN_BACSP	P29019 bacillus sp
12	91	8.1	578	1 LIPA_BACPU	O50774 mycoplasma
13	90.5	8.1	1169	1 CIGB_BACTZ	O92426 bacillus th
14	89.5	8.0	416	1 ELT1_CAEEL	P28515 caenorhabdi
15	89	7.9	599	1 LAC2_THACU	O02075 tharaphor
16	89	7.9	858	1 CYAG_DICDI	O03101 dictyostell
17	88.5	7.9	520	1 AMT_BACME	P20845 bacillus me
18	88.5	7.9	717	1 UBPB_YEAST	P36026 saccharomyc
19	87.5	7.8	382	1 Y077_METUA	O60384 methanococc
20	87.5	7.8	523	1 RPN3_YEAST	P40016 saccharomyc
21	87.5	7.8	626	1 PC11_YEAST	P39681 saccharomyc
22	87.5	7.8	1251	1 RBR2_PLAVB	O00799 plasmodium
23	87.5	7.8	1435	1 DPO3_MYCPU	P47729 mycoplasma
24	87	7.8	668	1 PBPC_BACSU	P42971 bacillus su
25	87	7.8	762	1 SLAP_ACEKI	P22258 acetogenium
26	87	7.8	1433	1 SUBE_BACSU	P16397 bacillus su
27	86.5	7.7	475	1 Y123_MYCPN	P75513 mycoplasma
28	86.5	7.7	678	1 YNC7_YEAST	P53368 saccharomyc
29	86.5	7.7	1694	1 IGA0_HAEIN	P44369 haemophilus
30	86.5	7.7	1702	1 IGA2_HAEIN	P45884 haemophilus
31	86	7.7	437	1 ACMA_LACLC	O48603 lactococcus
32	86	7.7	617	1 Y237_BUCAI	P57331 buchnera ap
33	86	7.7	986	1 GUNZ_CLOSR	P23659 clostridium

34	85.5	7.6	938	1 PM15_CHLPN	O92883 chlamydia p
35	85.5	7.6	1146	1 AS10_YEAST	P48361 saccharomyc
36	85.5	7.6	1277	1 IF3X_YEAST	O03690 saccharomyc
37	85	7.6	691	1 P21_YEAST	P26570 saccharomyc
38	85	7.6	1088	1 VAPA_CANTR	P38078 candida tiro
39	84.5	7.5	415	1 EP1G_YEAST	P29547 saccharomyc
40	84.5	7.5	925	1 PDC2_YEAST	P32896 saccharomyc
41	84.5	7.5	1569	1 YPJA_ECOLI	P52143 escherichia
42	84.5	7.5	1682	1 MSP1_PLAF3	P19598 plasmodium
43	84.5	7.5	2652	1 RRPB_IBVB	P26314 avian infec
44	84	7.5	401	1 FLGE_ECOLI	P75937 escherichia
45	84	7.5	437	1 YABE_BACSU	P37546 bacillus su

ALIGNMENTS

RESULT 1
ID CNA_STAU STANDARD: PRT: 1183 AA.
AC 053654;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE COLLAGEN ADHESIN PRECURSOR.
GN CNA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PDA 574;
RX MEDLINE=92165839; PubMed=1311320;
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wlberg K.,
RA Lindberg M., Hoeek M.;
RT Molecular characterization and expression of a gene encoding a
RT Staphylococcus aureus collagen adhesin."
RL J. Biol. Chem. 267:4766-4772(1992).
RN [2]
RN ERRATUM.
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wlberg K.,
RA Lindberg M., Hoeek M.;
RL J. Biol. Chem. 269:11672-11672(1994).
RN [3]
RN COLLAGEN-BINDING DOMAIN.
RC STRAIN-PDA 574;
RX MEDLINE=94032261; PubMed=8218209;
RA Patti J.M., Boles J.O., Hoeek M.;
RT Identification and biochemical characterization of the ligand
RT binding domain of the collagen adhesin from Staphylococcus aureus."
RL Biochemistry 32:11428-11435(1993).
RN [4]
RN X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
RX MEDLINE=97475225; PubMed=9334749;
RA Smyersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
RA Moore D., Jin L., Schneider A., Delucas L.J., Hoeek M.,
RA Narayana S.V.L.;
RT "Structure of the collagen-binding domain from a Staphylococcus
RT aureus adhesin."
RL Nat. Struct. Biol. 4:833-838(1997).
CC -1- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO
CC COLLAGEN-CONTAINING SUBSTRATA.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. CELL WALL.
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch).

CC EMBL; M81736; AAA20874.1; .

DR PDB; 1AMX; 24-JUN-98.

DR InterPro; IPR001899; Gram_pos_anchor.

DR PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.

KW Signal; Repeat; Transmembrane; Cell wall; 3D-structure.

FT SIGNAL 1 29

FT CHAIN 30 1183

FT DOMAIN 30 1157

FT TRANSMEM 1158 1177

FT DOMAIN 1178 1183

FT DOMAIN 151 318

FT DOMAIN 533 1093

FT DOMAIN 1093 1157

FT DOMAIN 1151 1156

FT REPEAT 533 719

FT REPEAT 720 906

FT REPEAT 907 1093

FT REPEAT 1183 AA; 133066 MW; B6A1CC072E575D76 CRC64;

SEQUENCE

Query Match 93.4%; Score 1048; DB 1; Length 1183;

Best Local Similarity 100.0%; Pred. No. 1,le-69;

Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 SDDKATITSGNKSTVWTVHKSSEAGTSVFFYKKTGDMLEPDTTHVWFLNINNEKSYVSK 71

DB 144 SDDKATITSGNKSTVWTVHKSSEAGTSVFFYKKTGDMLEPDTTHVWFLNINNEKSYVSK 203

QY 72 DITKDOIGGQOOLDLSTLNINVTGTHSNVYSGQSAITDEKAFPSKTIIVDTNKTIDV 131

DB 204 DITKDOIGGQOOLDLSTLNINVTGTHSNVYSGQSAITDEKAFPSKTIIVDTNKTIDV 263

QY 132 TTPQGGYSGNSNSINKTKITNEQOKEFPVNSQAWQOEKGKEVNSKSNHYHNINANA 191

DB 264 TTPQGGYSGNSNSINKTKITNEQOKEFPVNSQAWQOEKGKEVNSKSNHYHNINANA 323

QY 192 GIEGYKGLKYLKODKDKR 211

DB 324 GIEGYKGLKYLKODKDKR 343

RESULT 2

HSEB_CAVPO STANDARD: PRT: 1076 AA.

ID HSEB_CAVPO P70106;

DT 01-NOV-1997 (rel. 35, Created)

DT 01-NOV-1997 (rel. 35, Last sequence update)

DT 20-AUG-2001 (rel. 40, Last annotation update)

DE HEAT-STABLE ENTEROTOXIN RECEPTOR PRECURSOR (GC-C) (INTESTINAL GUANYLATE CYCLASE) (EC 4.6.1.2) (STA RECEPTOR) (GUANYLYL CYCLASE C).

OS GUCY2C OR GUCYC.

OS Cavia porcellus (Guinea pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Hystriocognathi; Cavidae; Cavia.

OX NCBI_TaxID=10141;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Colon;

RA Kruhoefter W., Cetin Y., Kaempf U., Forssmann W.-G.;

RL Submitted (Aug-1996) to the EMBL/Genbank/DBJ databases.

CC -1- FUNCTION: RECEPTOR FOR THE E. COLI HEAT-STABLE ENTEROTOXIN (E. COLI ENTEROTOXIN MARKEDLY STIMULATES THE ACCUMULATION OF GMP IN MAMMALIAN CELLS EXPRESSING GC-C). ALSO ACTIVATED BY THE ENDOGENOUS PEPTIDE GUANYLIN.

CC -1- CATALYTIC ACTIVITY: GTP = 3, '5'-CYCLIC GMP + PYROPHOSPHATE.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL CYCLASE FAMILY.

CC -1- SIMILARITY: CONTAINS 1 PROTEIN KINASE-LIKE DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; Z74734; CAA98989.1; .

DR HSSP; 002846; IAWL.

DR InterPro; IPR001828; ANF_receptor.

DR InterPro; IPR00719; Euk_pkinase.

DR InterPro; IPR001054; Guanylyl_cyclase.

DR Pfam; PF01094; ANF_receptor; 1.

DR Pfam; PF00211; guanylate_cyc; 1.

DR Pfam; PF00069; pkinase; 2.

DR SMART; SM00044; CYCC; 1.

DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.

DR PROSITE; PS00125; GUANYLATE_CYCLASES_2; 1.

DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.

KW Receptor; Transmembrane; Glycoprotein; Lyase; GMP synthesis; Signal.

FT SIGNAL 1 23

FT CHAIN 24 1076

FT DOMAIN 24 433

FT TRANSMEM 434 457

FT DOMAIN 458 1076

FT DOMAIN 492 752

FT DOMAIN 827 957

FT CARBOHYD 35 35

FT CARBOHYD 82 82

FT CARBOHYD 191 191

FT CARBOHYD 191 191

FT CARBOHYD 198 198

FT CARBOHYD 287 287

FT CARBOHYD 306 306

FT CARBOHYD 310 310

FT CARBOHYD 348 348

FT CARBOHYD 360 360

FT CARBOHYD 405 405

SO SEQUENCE 1076 AA; 123119 MW; 9B53F16B5E80EB9 CRC64;

Query Match 9.4%; Score 106; DB 1; Length 1076;

Best Local Similarity 19.6%; Pred. No. 2;

Matches 47; Conservative 37; Mismatches 78; Indels 78; Gaps 10;

QY 24 KSTNVTVHKSSEAGTSVFFYKKTGDMLEPDTTHVWFLN-INNEKSYVSKDITKDOIGG 82

DB 179 KASNLPRKFSFNTSYV--KNG-----TSEHCFFVIMNLEAGVSYFSQVGLGKEMLRGN 232

QY 83 QOILD-----LSTLNINVTGTH-----SNVYSGQSAITDF 111

DB 233 EELQKILKDPNRRSNIVNMGFPQWESLKIOWTATEDVYIILVDLFNNVYLEENATAP- 291

QY 112 EKAFPSKTIIVDTNKTIDVITTPQGGYSGNSPSINKTKITNEQOKEFPVNSQAW----- 166

DB 292 -----DYMKLVVLTLPNGSTIN-----TSLSKESLQEFDFALAYLDGL 333

QY 167 -----YOEKGKEVNSKSNHYHNINANA-----EGYKGLKYLKODKDKR 211

DB 334 LGCHMLKTLRLNG-ENTTAHKAFAHARNLTFEGSTGPVTLDDSGDIDNRMVLLVTSVDIK 392

RESULT 3

ATL_STAUP STANDARD: PRT: 1256 AA.

ID ATL_STAUP P52081;

DT 01-OCT-1996 (rel. 34, Created)

DT 01-OCT-1996 (rel. 34, Last sequence update)

DT 20-AUG-2001 (rel. 40, Last annotation update)

DE BIFUNCTIONAL AUTOLYSIN PRECURSOR (INCLUDES: N-ACETYLORAMOYL-L-ALANINE AMIDASE (EC 3.5.1.28); MANNOSYL-GLYCOPROTEIN ENDO-BETA-N-ACETYLGLUCOSAMIDASE (EC 3.2.1.96)).

GN ATL.

OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 OX NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 205-214 AND 776-792.
 RC STRAIN-FN450;
 RX MEDLINE=95116542; PubMed=7816834;
 RA Oshida T., Sugai M., Komatsuzawa H., Hong Y.-M., Sugihara H.,
 RA Tomasz A.;
 RT "A Staphylococcus aureus autolysin that has an N-acetylmuramoyl-L-
 RT alanine amidase domain and an endo-beta-N-acetylglucosaminidase
 RT domain: cloning, sequence analysis, and characterization.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:285-289(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MCTC 8325-4;
 RA Foster S.J.;
 RL Submitted (Apr-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: ENDOLYOLYSIS OF THE DI-N-ACETYLCHITOSYL UNIT IN
 CC HIGH-MANNOSE GLYCOPROTEINS AND GLYCOPROTEINS CONTAINING THE
 CC -[MAN)5(GLCNAc)2]-ASN STRUCTURE. ONE N-ACETYL-D-GLUCOSAMINE
 CC RESIDUE REMAINS ATTACHED TO THE PROTEIN; THE REST OF THE
 CC OLIGOSACCHARIDE IS RELEASED INTACT.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSES THE LINK BETWEEN N-ACETYLMURAMOYL
 CC RESIDUES AND L-AMINO ACID RESIDUES IN CERTAIN BACTERIAL CELL-WALL
 CC GLYCOPROTEINS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- PTM: UNDERGOES PROTEOLYTIC PROCESSING TO GENERATE THE TWO
 CC EXTRACELLULAR LYTIC ENZYMES.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE N-
 CC ACETYLMURAMOYL-L-ALANINE AMIDASE FAMILY 2.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 73 OF
 CC GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: D17366; BAA04185.1; -
 CC EMBL: L41499; AAA89982.1; -
 CC InterPro: IPR002502; Amidase_2.
 CC InterPro: IPR002901; Amidase_4.
 CC Pfam: PF01510; Amidase_2; 1.
 CC Pfam: PF01832; Amidase_4; 1.
 CC SMART: SM00047; LY22; 1.
 CC Cell wall; Hydrolyase; Signal; Multifunctional enzyme; Repeat.
 CC KW SIGNAL; 1
 CC FT CHAIN 1 29
 CC FT DOMAIN 30 1256 BIFUNCTIONAL AUTOCLYSIN.
 CC FT DOMAIN 199 775 N-ACETYLMURAMOYL-L-ALANINE AMIDASE.
 CC FT DOMAIN 776 1256 ENDO-BETA-N-ACETYLGLUCOSAMIDASE.
 CC FT REPEAT 425 589
 CC FT REPEAT 589 758
 CC FT REPEAT 758 932
 CC FT REPEAT 932 1256
 CC SEQUENCE 1256 AA; 137384 MW; 2BB76CAA292FDD20 CRC64;
 SQ
 Query Match 8.7%; Score 98; DB 1; Length 1256;
 Best Local Similarity 21.6%; Pred. No. 9.5;
 Matches 61; Conservative 39; Mismatches 90; Indels 92; Gaps 16;
 Oy 15 KYATINSGKSTVYHKEAGTSVFE---YKTKGD-----MLPEPTTH---VRNF 59
 Db 790 KTAQVPRNNTGIRASYETKAKGAKYADRTYVTERKAGNTYLLNNTSINIFLGWF 849
 Oy 60 ---ININN-EKSYVSKDITKDOIQGOQDLSTLN-INVTGHSNYSYSGSATDFE 112
 Db 850 NVADLNVQNLGKREVKTQYKYVTKSNNGLSMVPWGTKNQVILTGN---IAQGTNARK 905

Oy 113 KAPGSKI---TVDNTK---NTIDVTIP-----OGYGSY-----N 141
 Db 906 QVSVGADVLYXGRIINNRGTGVNNAKDLAPYAVKPTISAADVYTYVKNNGNYVTPN 965
 Oy 142 SFSINRYKTKITNEQO---KEFYNSQAWYOEHGKEEVNCK-----SFNHT 183
 Db 966 SDTAKYSLKAFNEQPPAVVVEQVINGQTYWY--YGKLS-NGKLMKSTDLAKELIKYNOT 1022
 Oy 184 VHNINANAGIEGTVKELKY-----LQDKRTK 211
 Db 1023 GMTLNOVAQIQAGLYQKPPQVRVPGKWTDAKFNVDXHAMDTK 1064
 RESULT 4
 TOP1_BORBU STANDARD; PRT; 848 AA.
 ID TOP1_BORBU
 AC 051768;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DNA TOPOISOMERASE I (EC 5.99.1.2) (OMEGA-PROTEIN) (RELAXING ENZYME)
 DE (UNWISTING ENZYME) (SWIHELASE).
 GN TOPA OR BB0828.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=9403685;
 RA Fraser C.M., Castjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwin M.,
 RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
 RA Peterson J., Kierlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
 RA Van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
 RA Utechtback T., Matthey L., McDonald L., Artlich P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT Genomic sequence of a Lyme disease spirochete, Borrelia
 RT burgdorferi.";
 RL Nature 390:580-586(1997).
 CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
 CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.
 CC -1- CATALYTIC ACTIVITY: ATP-INDEPENDENT BREAKAGE OF SINGLE-STRANDED
 CC DNA, FOLLOWED BY PASSAGE AND REJOINING.
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
 CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
 CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
 CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.
 CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
 CC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: AE001180; AAC67161.1; -
 CC TIGR: BB0828; -
 CC InterPro: IPR003601; DNATopI_ATP_bind.
 CC InterPro: IPR003602; DNATopI_DNA_bind.
 CC InterPro: IPR000380; Pro_topoisomase.
 CC InterPro: IPR002936; Toprim.
 CC Pfam: PF01131; Topoisom_Dac; 1.
 CC Pfam: PF01751; Toprim; 1.
 CC PRINTS: PR00417; PRPISMRASEI.
 CC SMART: SM00437; TOP1AC; 1.
 CC SMART: SM00436; TOP1BC; 1.
 CC SMART: SM00493; TOP1RM; 1.

DR PROSITE; PS00396; TOPOISOMERASE_I_PROK; 1.
 KW Isomerase; Topoisomerase; DNA-binding; Complete proteome.
 FT ACT_SITE 324 324 DNA_CLEAVAGE (BY SIMILARITY).
 SQ SEQUENCE 848 AA; 97663 MW; 37DEA5FB1E12FA2C CRC64;

Query Match 8.6%; Score 96.5; DB 1; Length 848;
 Best Local Similarity 22.8%; Pred. No. 7.5;
 Matches 59; Conservative 41; Mismatches 84; Indels 75; Gaps 14;

19 ITSGKSTNVTVHSEACTSVFYVY-----KTGDMLPEDTTHRWFLINNEKSY 68
 Db 457 IKKGDTEIVAKKTSSEHTKAPFRTEASLYQKMEKIGRPSTYSTI---ISTLLREY 513
 QY 69 VSK-DITIKDIOGQOQDL-----STLININVTGTH-----SNYS 103
 Db 514 AFKLNTLPTIKGAAYVLEKPEVPLIELNFTSNMEKLDKAIKGLDKIKYLSKFTYN 573
 QY 104 GOSATD-----FEKAPGSGKITVDNTKNTIDVTIPQG-YGSY-----NSFS 144
 Db 574 GKKGKLDVTVMLEPKIDSEFRVYIESOKI---ENKNSINVTINIKYGPYLIFKGNYS 630
 QY 145 INYKTKINEOQKE---FVNNSQAWYEHGKEEYNGKS--FNHTVH-NI-----NANAG 192
 Db 631 INAKTPELNLYKKDIEKTIINEKEIKPNILGVDPULGLNVIFKNTIYQVQLEGDTNAP 690
 QY 193 IEGTVKG---ELKVLKODK 208
 Db 691 QEYTKKGKPKLKIKAKK 709

RESULT 5
 MSN2_YEAST
 ID MSN2_YEAST STANDARD; PRT; 704 AA.

DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE ZINC FINGER PROTEIN MSN2 (MULTICOPY SUPPRESSOR OF SNF1 PROTEIN 2).
 GN MSN2 OR YMR037C OR YMS932.02C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;

RA SEQUENCE FROM N.A.
 RP STRAIN=MS288C;
 RC MEDLINE=93309420; PubMed=83211194;
 RX Estruch F., Carlson M.;
 RA "Two homologous zinc finger genes identified by multicopy suppression
 in a SNF1 protein kinase mutant of Saccharomyces cerevisiae.";
 RT Mol. Cell. Biol. 13:3872-3881(1993).
 RL [2]
 RN RP SEQUENCE FROM N.A.
 RC STRAIN=MS288C / AB972;
 RA Odeil C., Bowman S., Barrell B.G., Rajadream M.A.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=96208509; PubMed=8641288;
 RA Martinez-Pastor M.T., Marchler G., Schnell C., Marchler-Bauer A.,
 RA Ruls H., Estruch F.;
 RT "The Saccharomyces cerevisiae zinc finger proteins Men2p and Men4p
 are required for transcriptional induction through the stress
 response element (SRE).";
 RT EMO J. 15:2227-2235(1996).
 RL [1]
 CC -1- FUNCTION: POSITIVE TRANSCRIPTIONAL FACTOR THAT ACTS AS A COMPONENT
 OF THE STRESS RESPONSIVE SYSTEM. RECOGNIZES AND BINDS TO THE
 STRESS RESPONSE ELEMENT (SRE) WHICH IS INVOLVED IN THE RESPONSE
 TO VARIOUS FORMS OF STRESS (HEAT, OXIDATIVE, OSMOTIC, ETC.).
 CC INVOLVED IN THE REGULATION OF THE CTT1, DDR2, HSP12 GENES.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: TO MAMMALIAN EGR PROTEINS IN THE ZINC FINGER

CC REGIONS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; L08838; AAA34806.1; -
 DR EMBL; Z48502; CAAB8403.1; -
 DR PIR; A48131; A48131.
 DR PIR; S39004; S39004.
 DR HSSP; P08047; 1SP1.
 DR TRANSFAC; T01257; -
 DR SGD; S0004640; MSN2.
 DR InterPro; IPR000822; Znf-C2H2.
 DR Pfam; PF00096; Zf-C2H2; 2.
 DR PRINTS; PR00048; ZINC_FINGER.
 DR SMART; SM00355; Znf_C2H2; 2.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 2.
 KW DNA-binding; Nuclear protein; zinc finger; Metal-binding; Activator;
 KW Transcription regulation.
 FT DOMAIN 1 109 ASP-RICH (ACIDIC).
 FT DOMAIN 260 279 ASP-RICH (ACIDIC).
 FT ZN_FING 647 665 C2H2-TYPE.
 FT ZN_FING 676 698 C2H2-TYPE.
 SQ SEQUENCE 704 AA; 77860 MW; EDF6F07446819DF1 CRC64;

Query Match 8.6%; Score 96; DB 1; Length 704;
 Best Local Similarity 25.4%; Pred. No. 6.5;
 Matches 64; Conservative 31; Mismatches 97; Indels 60; Gaps 13;

QY 12 SDDKVAITTSKSKSTNVVHSEACTSVFYIKTGDMLPEDTTHRWFLINNEKSYSK 71
 Db 229 SDTNSYSINSISNSNSNGNLN---SSYFNSLNDISMLDVSDDLINDDDTNTLSR 284
 QY 72 ---DITIKDIOG-----GOQLDL-----STLININVTGTHSNYSQOSAT 109
 Db 285 RRFSSV-ITNQPSPMTNSRNSISLSDLMNHPKIPSNRNTMLNT-INSISSNASPRT 342
 QY 110 DEKAPGSGKITVDNTKN---TIDVTIPQGYGYN-----SFSI 145
 Db 343 TTMANADSNL-AGNPKNNDATIDNELQILNEYMNFNNDLGTSGKNKACPSFDA 401
 QY 146 NYKTIITEQO-KEFVNNSQ--AMYOEGKEEYNGKSRNHYHINANAG---IEGYKG 199
 Db 402 NAMTINPSOOLQOOLNRYOHKOLTSNHNSSSTNMKSFNSDLRSRORASLPIDDSLSY 461
 QY 200 ELKVLKODKDFK 211
 Db 462 DL-VNKODEDPK 472

RESULT 6

MUR2_ENTHR
 ID MUR2_ENTHR STANDARD; PRT; 666 AA.

DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MURAMIDASE-2 PRECURSOR (EC 3.2.1.17) (1,4-BETA-N-
 ACETYL-MURAMOTYLHYDROLASE) (PEPTIDOLYCAN HYDROLASE) (PG-HYDROLASE-2)
 DE (LISOSYME).
 OS Enterococcus hirae.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
 CC Enterococcus.
 OX NCBI_TaxID=1354;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 50-73.

CC STRAIN-ATCC 9790;
RX MEDLINE-92165737; PubMed-1347040;
RA Chu C.-P., Kariyama R., Daneo-Moore L., Shockman G.D.;
RT "Cloning and sequence analysis of the muramidase-2 gene from
RT *Enterococcus hirae*."; [2]
RL J. Bacteriol. 174:1619-1625(1992).
RN [2]
RP FUNCTION
RC STRAIN-ATCC 9790;
RX MEDLINE-89327152; PubMed-2753858;
RA Dolinger D.L., Daneo-Moore L., Shockman G.D.;
RT "The second peptidoglycan hydrolase of *Streptococcus faecium* ATCC 9790
RT covalently binds penicillin";
RL J. Bacteriol. 171:4355-4361(1989).
RN [3]
RP FUNCTION
RC STRAIN-ATCC 9790;
RX Del Mar Lleo M., Canepari P., Satta G.;
RT "Thermosensitive cell growth mutants of *Enterococcus hirae* that
RT elongate at non-permissive temperature are stimulated to divide by
RT parental autolytic enzymes.";
RL J. Gen. Microbiol. 139:3099-3117(1993).
CC -1- FUNCTION: MAY WORK IN CONCERT WITH AND POTENTIATE THE PROCESSIVE
CC HYDROLYTIC ACTION OF MURAMIDASE-1, WHICH REQUIRES BINDING OF THE
CC ENZYME TO NONREDUCING ENDS OF GLYCAN CHAINS. HYDROLYSIS IN THE
CC MIDST OF GLYCAN CHAINS WOULD INCREASE THE NUMBER OF BINDING SITES
CC FOR MURAMIDASE-1. MAY FUNCTION IN FACILITATING SEPTUM FORMATION
CC AND CELL SEPARATION. ACTIVE ON M.LTREUS CELL WALLS AND ON E.HIRAE
CC CELL WALL FRACTIONS, BUT NOT ACTIVE ON E.HIRAE INTACT CELL WALLS.
CC CAN COVALENTLY BIND PENICILLIN.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES BETWEEN N-
CC ACETYL-D-GLUCOSAMINE AND N-ACETYLIMRAMIC ACID IN PEPTIDOGLYCAN
CC HETEROPOLYMERS OF THE PROKARYOTES CELL WALLS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN
CC BINDING.
CC -1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
CC -1- SIMILARITY: CONTAINS 6 LYSM REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M77639; AAA24776.1; -
DR PIR: A42296; A42296.
DR HSSP: P21697; 1PCS.
DR Interpro: IPR002901; Amidase_4.
DR Interpro: IPR002482; lysm.
DR Pfam: PF01832; Amidase_4; 1.
DR Pfam: PF01476; Lysm; 6.
DR SMART: SM00257; Lysm; 6.
DR SMART: SM00047; Lys2; 1.
DR Hydrolase: Glycosidase; Bacteriolytic enzyme; Cell wall;
KW Cell division; Septation; Repeat; Signal.
FT SIGNAL 1 49
FT CHAIN 50 666 MURAMIDASE-2.
FT REPEAT 257 299 LYSM 1.
FT REPEAT 338 380 LYSM 2.
FT REPEAT 414 456 LYSM 3.
FT REPEAT 489 531 LYSM 4.
FT REPEAT 565 607 LYSM 5.
FT REPEAT 623 665 LYSM 6.
FT REPEAT 666 70670 MW; FFOAA/FAFCD810BA3 CRC64;
SQ SEQUENCE 666 AA; 70670 MW; FFOAA/FAFCD810BA3 CRC64;

Query Match	8.5%;	Score 95;	DB 1;	Length 666;
Best Local Similarity	24.0%;	Pred. NO. 7.2;		
Matches 49;	Conservative 28;	Mismatches 81;	Indels 46;	Gaps 11;

```

OY 20 TSGKSTVTVYHKSEAGSSVVEYKKTGMDLPEDTTH-----VRMFLNNKESKYS 70
Db 242 TSGNSGSGSAT--TGTITYVKSQSV-WGISHSGFCTMAOLIEV-NNIKNNFTPG 293
OY 71 KDIIRKDDIOGGQOOLDLSTLNI-----NVTG-THSNYSGOSATIDFEKAPGSKITV 122
Db 294 QKLTIK-----CGGASGSSTTMTGNNASGNGTSGNTSGTSGQAT-----GAKTV 340
OY 123 DNTKNTIDVTIPGCGYSYNSFSINKYKTIINEOQKEFYNNSSQAWOEHKEEYKGSFNH 182
Db 341 KSGDSVMKIANDHG-----ISMNOLIE-WNNIKNNFEPYQOOLVWSSKSSSAGSTSTNT 393
OY 183 TVHNNANAGIEG-TVGKELKVK 205
Db 394 STGNTSSNTANTGTTSGSTYTVK 417

RESULT 7
YD_BA_ECOLI STANDARD; PRT: 2003 AA.
ID YD_BA_ECOLI
AC P31666; P76087; P76088; P76856; P76857; P76859;
DT 01-FEB-1994 (Rel. 28, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHEICAL 205.9 KDA PROTEIN IN PAAY-ACPD INTERGENIC REGION.
GN YD_BA OR B1401/B1405.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
CX NCBI_TaxID=562;
OY [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio H., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiiuchi T.;
RA "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RP [3]
RP SEQUENCE OF 464-2003 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92190338; PubMed=1665988;
RA Moszer I., Glasner P., Danchin A.;
RT "Multiple IS insertion sequences near the replication terminus in
RL Escherichia coli K-12.";
RL Biochimie 73:1361-1374(1991).
CC -1- SIMILARITY: TO S.TYPHIMURUM ORF NEAR CYSG (AC P25928).
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR
CC THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT
CC BETWEEN AMINO ACIDS 839 AND 840.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch.)

```

CC EMBL; AE000237; AAC74483.1; ALT_SEQ.
 DR EMBL; AE000237; AAC74483.1; ALT_SEQ.
 DR EMBL; D90778; BAA15009.1; ALT_SEQ.
 DR EMBL; D90778; BAA18880.1; ALT_SEQ.
 DR EMBL; D90779; BAA18881.1; ALT_SEQ.
 DR EMBL; X62680; -; NOT_ANNOTATED_CDS.
 DR Ecogen; B61307; ydaA.
 KM Hypothetical protein; Complete proteome.
 FT CONFLICT 489 489 I -> V (IN REF. 2).
 FT CONFLICT 495 495 I -> V (IN REF. 2).
 SO SEQUENCE 2003 AA; 205949 MW; B83A12C8B5220BE CRC64;

Query Match 8.5%; Score 95; DB 1; Length 2003;
 Best Local Similarity 25.1%; Pred. No. 28;
 Matches 69; Conservative 30; Mismatches 92; Indels 84; Gaps 16;
 QY 5 HHHHHG-----SDDKAVTISGK-----STNVYVHKSQA 35
 DB 94 HHHRRNSPLPPPPDESDTVPVPPGDEIIPDDPDPVPPKPVSPNNVYIDKTEK 153
 QY 36 GTS-----SVFY---KTGDLPEDTTHVRFELNI-----NNEKSY--VSKDITIKDIOG 81
 DB 154 TITFINDSVFTYENADGTISLDSNGRKATINLMQIDEANNVALEGVASADGATKQVNH 213
 QY 82 GOOL-----DLSTLNIN-----VTGTHSNVYSGQSAITPEK-----AFPGSK 119
 DB 214 NEELVYTGDNATVGNNGKTVGDKDSTGEINGNNGK-VIQQGDLDSGGHIDITGS 272
 QY 120 ITVDNKTNTIDVTIPQGYG---SYNSFSINT--KTKITNEQKEFYNNQAOYOEKKE 174
 DB 273 ATVDN-KGTMTVTWDPESMGIDGDKAIVNNEGSEFTTGTGTQINGDADATANNNGKTT 331
 QY 175 VNGKSEPHVHNINANAGIEGTVKELKYLKODK 209
 DB 332 VDGKDSGT--EINGNNG-----KVI-ODGD 354

RESULT 8
 BXG_CLOBO STANDARD; PRT: 1296 AA.
 AC 060393;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE BOTULINUM NEUROTOXIN TYPE G PRECURSOR (EC 3.4.24.69) (BONT/G)
 GN BONT.
 OS Clostridium botulinum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=113 / 30;
 RX MEDLINE=94092745; PubMed=8268233;
 RA Campbell K., Collins M.D., East A.K.;
 RT "Nucleotide sequence of the gene coding for Clostridium botulinum
 (Clostridium argentinense) type G neurotoxin: genealogical comparison
 with other clostridial neurotoxins."
 RL Biochim. Biophys. Acta 1216:487-491(1993).
 CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
 AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
 WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 ENDOPEPTIDASE.
 CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
 HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
 WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
 FORMATION AND TOXIN BINDING, RESPECTIVELY.
 CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).

CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
 CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
 CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
 CC SUPERFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC
 DR EMBL; X74162; CAA52275.1; -.
 DR HSSP; P04958; IAF9.
 DR MEROPS; M27.002; -.
 DR InterPro; IPR000395; Bontoxilysin.
 DR InterPro; IPR000130; Zn_Mrpeptidase.
 DR Pfam; PF01742; Peptidase_M27; 1.
 DR PRINTS; PR00760; BONTOXILYSIN.
 DR Prodom; PD001963; Bontoxilysin; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KM Neurotoxin; Hydrolase; Metalloprotease; Zinc.
 FT INIT_MET 0
 FT CHAIN 1 441
 FT CHAIN 442 1296
 FT METAL 229 229
 FT ACT_SITE 230 230
 FT METAL 233 233
 FT DISULFID 435 449
 SO SEQUENCE 1296 AA; 149013 MW; DC8E47E15F665C1 CRC64;

Query Match 8.3%; Score 93.5; DB 1; Length 1296;
 Best Local Similarity 23.0%; Pred. No. 21;
 Matches 50; Conservative 32; Mismatches 66; Indels 69; Gaps 12;
 QY 47 DMLP-----EDTHVRFELNINNEKSYVSKDITIKDIOGGQOOLDS---TLNINVT 95
 DB 838 DSIFPDLISLYTKDTILIOYF---NNYISMISSAIIISLYRGRRLDSSGYGATMNVGSD 894
 QY 96 GTHSNVYSGQSAITPEKAPGSKITVDNKTNTIDVTIPQGYGS-YNSFSINT----- 147
 DB 895 VIFNDIGNQOFKLNNSE---NSNITAHQSKFYV-----YDSMFDNFSINFWATPKY 943
 QY 148 -KTKITNEQKEF-----VNNQSAW-YQEHKE-----EVNGKS----- 179
 DB 944 NNNDIQFYLDNEYTIIISCKNDGKMSVSIKGNRIITWLLIDVNAKSKSIPEYSIKDNISD 1003
 QY 180 -----FNHTVHNIN--NANAGIEGTVKELKYLKODK 208
 DB 1004 YINKWFSTITTDRLGNANIYINGSLAKSEKILINDR 1040
 RESULT 9
 TOP1_MYCPN STANDARD; PRT: 711 AA.
 AC P78032;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DNA TOPOISOMERASE I (EC 5.9.9.1.2) (OMEGA-PROTEIN) (RELAXING ENZYME)
 GN (UNWISTING ENZYME) (SWIVELASE).
 GN TOPA OR MRN261 OR MP572.
 OS Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 CC Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 29342 / M129;
 RX MEDLINE=9710585; PubMed=8948633;

Db 131 FKPYIGARVAVGHVHRHSIDSTKKTIEVTYPSNAPNGAVTTNTDPKT-----QNDYOS 184

QY 162 NS 163

Db 185 NS 186

RESULT 11

GUN_BACSP STANDARD; PRT; 463 AA.

ID GUN_BACSP

AC P29019;

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULOSE) (ENDO-K).

OS Bacillus sp. (strain KSM-330).

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxId=1409;

RP SEQUENCE FROM N.A.

RX MEDLINE=92121880; PubMed=1770347;

RA Ozaki K., Sumitomo N., Ito S.;

RT "Molecular cloning and nucleotide sequence of the gene encoding an endo-1,4-beta-glucanase from Bacillus sp. KSM-330.";

RL J. Gen. Microbiol. 137:2299-2305(1991).

RN [2]

RP SEQUENCE OF 56-75, AND CHARACTERIZATION.

RX MEDLINE=91259037; PubMed=2045781;

RA Ozaki K., Ito S.;

RT "Purification and properties of an acid endo-1,4-beta-glucanase from Bacillus sp. KSM-330.";

RL J. Gen. Microbiol. 137:41-48(1991).

CC -1- FUNCTION: THIS ACID ENDOGLUCANASE IS ACTIVE OVER AN EXTREMELY NARROW RANGE OF PH VALUES, BETWEEN 4.5 AND 6.5, WITH AN OPTIMUM PH AT 5.2.

CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.

CC -1- PTM: THE N- AND THE C-TERMINUS MAY BE SUBJECTED TO PROTEOLYSIS.

CC -1- MISCELLANEOUS: ONE TRP RESIDUE HAS BEEN PROVED TO BE INVOLVED IN THE MECHANISM OF ACTION OF ENDO-K.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY D (FAMILY 8 OF GLYCOSYL HYDROLASES).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL: M68872; AAA22409.1; -.

CC PIR: A44808; A44808.

DR HSSP: P04955; ICHEM.

DR InterPro: IPR002037; Glyco_hydro.8.

DR Pfam: PF01270; Glyco_hydro.8; 1.

DR PRINTS: PRO0735; GLHYDRLASE8.

DR PROSITE: PS00812; GLYCOSYL_HYDROL_F8; 1.

CC Cellulose degradation; Hydrolase; Glycosidase; Signal.

FT SIGNAL 1

FT PROPEP 2

FT CHAIN 56

FT ACT_SITE 130

FT ACT_SITE 191

SO SEQUENCE 463 AA; 51882 MW; 407FA54F5236C59E CRC64;

Query Match 8.1%; Score 91; DB 1; Length 463;

Best Local Similarity 21.2%; Pred. No. 9;

Matches 53; Conservative 31; Mismatches 64; Indels 102; Gaps 13;

QY 7 HHHHGSDDKY-----ATTSGNKSTNYT-----VHKSEAGTSVSEFYKGTDMIPED--T 53

Db 204 HKQWSSGSKINTYKEAQNMTTKGKASVTKNNGNLNGDWGDSSTF-----DTRPSDMM 238

QY 54 TVHWFLINNNKSKSYVKDITIKQIGGOGLDLSTLNI--NTYGTSHNY---YSGOSA- 107

Db 259 SHLRAFYEFPTGDKTW-----LNVLDNLXNTYNTNTNKSPTKGL 297

QY 108 ITDF--EKAEPGSKITVDNTKNT-----IDVTIPQYGSYN-----SFS 144

Db 298 ISDFVVKNPPOAPKDBLDESKYTDSEYYNARSVPRLRVMDYAMYGEGKRGKISDKVATM 357

QY 145 INKTK-----ITNEQKEFVNNNSAHTQERG 171

Db 358 IKSRTKCNPSKITYDGYKLDGTNGDYPTAVYSPFIAGTTSKNQDWMVNSGMDW-MKNK 416

QY 172 KEENYKGSFN 181

Db 417 KEYSFSDSYN 426

RESULT 12

LIPAB_MYCPU STANDARD; PRT; 578 AA.

ID LIPAB_MYCPU

AC Q50274;

DT 20-AUG-2001 (Rel. 40, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE LIPOPROTEIN A PRECURSOR.

GN LIPAB OR MYPU_5300.

OS Mycoplasma pulmonis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;

OC Mycoplasmataceae; Mycoplasma.

OX NCBI_TaxId=2107;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=KD735-15;

RX MEDLINE=20245550; PubMed=10781561;

RA Shen X., Gumulak J., Yu H., French C.T., Zou N., Dybvig K.;

RT "Gene rearrangements in the *usa* locus of *Mycoplasma pulmonis*.";

RL J. Bacteriol. 182:2900-2908(2000).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=QAB CTIP;

RX MEDLINE=21267165; PubMed=11353084;

RA Chandraud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F., Moszer I., Dybvig K., Wroblewski H., Viart A., Rocha E.P.C., Blanchard A.;

RT "The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulmonis*.";

RL Nucleic Acids Res. 29:2145-2153(2001).

RN [3]

RP PARTIAL SEQUENCE FROM N.A.

RC STRAIN=KD735-15;

RX MEDLINE=96414471; PubMed=8817492;

RA Bhugra B., Voelker L.U., Zou N., Yu H., Dybvig K.;

RT "Mechanism of antigenic variation in *Mycoplasma pulmonis*: interwoven, site-specific DNA inversions.";

RL Mol. Microbiol. 18:703-714(1995).

CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID ANCHOR (POTENTIAL).

CC -1- SIMILARITY: BELONGS TO THE PULMONIS LIPAB LIPOPROTEIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL: U23947; AABA1030.2; -.

DR EMBL: AL445565; CAC13703.1; -.

DR Mypulisti: MYPU_5300: -
 KW Lipoprotein; Membrane; Signal; Complete proteome.
 FT SIGNAL 1
 FT CHAIN 27
 FT LIPID 28 578
 FT DOMAIN 82 165
 FT VARIANT 7 7
 FT VARIANT 41 41
 FT VARIANT 110 110
 FT VARIANT 127 127
 FT VARIANT 547 547
 SQ SEQUENCE 578 AA; 66219 MW; B533ED3467005D89 CRC64;
 Query Match 8.1%; Score 91; DB 1; Length 578;
 Best Local Similarity 20.2%; Pred. No. 12;
 Matches 49; Conservative 33; Mismatches 103; Indels 58; Gaps 6;
 QY 14 DKVATITSGNK-STNVYHAKSEAGTSVFYKGTGDMLEPETHVRWFLINNEKSYSKD 72
 DB 61 DKXNVSDNDKSTNKAVSNENSTOSOKTNESSONTKDDSSKTSNLTTONSSNTKSK 120
 QY 73 ITIKDIQGGQOLDLSTLINVTGTHSNYSGSAITDFEKAFFGSKITVDNTRKNTIDVT 132
 DB 121 IQENKQ-----SQKQNTSAVNSALE-----KQTKNDENISLVNSKDTNYILANDEKVA 170
 QY 133 IPQGTG---STNFSYNTKTKITNEQKEFYVNSQ---AWYOE----- 169
 DB 171 LAKDSREKSKSNLSNLTNKTPEVNROKNEVKDKALQWQKLNESASILESPSYDQTS 230
 QY 170 -----HGKEVNGKSFNHHVHINNAGIEGTAKELK 205
 DB 231 LSLTFKEGMPGLGLEVLVTKLEMLDHSHEKEISFKITNGKVNVLLTSSNLISGKMKIKSFS 290
 QY 206 QDK 208
 DB 291 FDK 293
 RESULT 13
 CIGB_BACT2 STANDARD; PRT; 1169 AA.
 AC 09ZAZ6;
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PESTICIDIAL CRYSTAL PROTEIN CRYIGB (INSECTICIDIAL DELTA-ENDOTOXIN
 DE CRYIG(B)) (CRYSTALLINE ENTOMOICIDAL PROTOXIN) (133 KDA CRYSTAL PROTEIN).
 GN CRYIGB OR CRYIG(B) OR CRYH2.
 OS Bacillus thuringiensis (subsp. wuhanensis).
 OC Bacillus; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_Taxid=52024;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HD-525;
 RX MEDLINE-20153386; PubMed-10688690;
 RA Kao W.-S., Lin J.-H., Tzeng C.-C., Kao S.-S., Chak K.-F.;
 RT "Cloning of two new cry genes from Bacillus thuringiensis subsp.
 RT wuhanensis strain";
 RT Curr. Microbiol. 40:227-232(2000).
 CC -1- FUNCTION: PROMOTES COLLOIDSMOTIC LYSIS BY BINDING TO THE MIDGUT
 CC EPITHELIAL CELLS OF LEPIDOPTERAN LARVAE. TOXIC TO PIERIS RAPAE.
 CC -1- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING
 CC SPOROGATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART
 CC OF THE SPORE COAT.
 CC -1- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE
 CC N-TERMINUS.
 CC -1- SIMILARITY: BELONGS TO THE DELTA-ENDOTOXIN FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U70725; AAD10291.1; -
 DR HSSP: P02965; ICIT.
 DR InterPro: IPR001178; Endotoxin.
 DR Pfam: PF00555; endotoxin; 1.
 KW Toxin; Sporulation.
 SQ SEQUENCE 1169 AA; 132904 MW; D1EFC1508A8B10BD CRC64;
 Query Match 8.1%; Score 90.5; DB 1; Length 1169;
 Best Local Similarity 24.5%; Pred. No. 31;
 Matches 36; Conservative 17; Mismatches 81; Indels 13; Gaps 5;
 QY 28 VYHKEAGTSVFYKGTGDMLEPETHVRWFLINNEKSYSKDITIKDIQGGQOLDL 87
 DB 1021 VTAKEGEGECVTHET-----EDNDELKFSNVEEITPNNTVTCNDTAAQEEIE- 1074
 QY 88 STLINVTGTHSNYSGSAITDFEKAFFGSKITVDNTRKNTIDVTIPQYGSYNSFSINY 147
 DB 1075 GTYTSRNNGYDGAVESNSVPADYASAVEKAYTGDGRDNCESN--RGYGDYTPPLPAGY 1132
 QY 148 KTKITNEQKEFYVNSQAWYOEHKKEE 174
 DB 1133 VTK-----ELEVEFETDKWMI-EIGETE 1154
 RESULT 14
 ETL1_CAEL
 ID ETL1_CAEL STANDARD; PRT; 416 AA.
 AC P28515;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TRANSCRIPTION FACTOR ETL-1.
 GN ETL-1 OR W09C2.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91342668; PubMed-1875944;
 RA Spieth J., Shim Y.H., Lea K., Conrad R., Bjumenthal T.;
 RT "elt-1, an embryonically expressed Caenorhabditis elegans gene
 RT homologous to the GATA transcription factor family.";
 RL Mol. Cell. Biol. 11:4651-4659(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Lloyd C.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF GENES CONTAINING THE GATA
 CC REGION, INCLUDING VITELLOGENIN GENES.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: CONTAINS 2 GATA-TYPE ZINC FINGERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X57834; CAA40967.1; -
 DR EMBL: Z68221; CAA92494.1; ALT_INT.
 DR PIR: A41267; A41267.
 DR HSSP: P17678; IGAU.
 DR WormPep: W09C2.1; CE03799.

DR InterPro: IPR000679; ZNF_GATA.
 DR Pfam: PF00320; GATA; 2.
 DR PRINTS: PR00619; GATAZNFINGER.
 DR SMART: SM00401; ZNF_GATA; 2.
 DR PROSITE: PS00344; GATA_ZN_FINGER_1; 2.
 DR PROSITE: PS00114; GATA_ZN_FINGER_2; 2.
 DR Transcription regulation; Activator; DNA-binding; Zinc-finger;
 KW Nuclear protein.
 FT ZN_FING 217 241 GATA-TYPE 1.
 FT ZN_FING 272 296 GATA-TYPE 2.
 SQ SEQUENCE 416 AA; 44823 MW; 01F5A19A57D42D CRC64;

Query Match 8.0%; Score 89.5; DB 1; Length 416;
 Best Local Similarity 22.0%; Pred. No. 10;
 Matches 45; Conservative 34; Mismatches 89; Indels 37; Gaps 9;

QY 11 GSDKATITSGKSNVYHKSSEAGTSVFYKTDMLPEDTHRMFINNENKSYVS 70
 DB 13 GISSGASLAPSTSTTASTAPPSYNTSATNNTNT-----TPSSYPMTNYO---YA 61
 QY 71 KDTIRKDIQGGQQLDLSTLINVTGTHSN--YYSQSATIDF-----EKAPFGSKITVD 123
 DB 62 GGTVTYTDMAFGSGMDMS-MNNGVFGQNNPSYFPTTOLNTYGYDTLAATTAAGTIVN 120
 QY 124 NIKNTIDVTIPGYSY--NSFSINKTKITNDQKEVNNNSQ-----AMVQEHKE 173
 DB 121 N-NOINVINVOGNGTIVPQITQNIITSVSNVQSSVPINNNSQPLPPTGLAGCSTSGSS 178
 QY 174 EVNGKSPN-----HTVHNINANAG 192
 DB 179 SASSSSANSTSPKNTISKANRSSG 203

RESULT 15
 LAC2_THACU STANDARD; PRT; 599 AA.
 AC 002075;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE LACCASE 2 PRECURSOR (EC 1.10.3.2) (BENZENEDIOL:OXYGEN OXIDOREDUCTASE)
 DE (URISHIOL OXIDASE) (DIPHENOL OXIDASE).
 GN LCC2.
 OS Thanaeophorus cucumeris (Black scurf of potato) (Rhizoctonia solani).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
 OC Heterobasidiomycetidae; Ceratobasidiales; Ceratobasidiaceae;
 OC mitosporic Ceratobasidiaceae; Rhizoctonia.
 OC NOBI_TaxID=107832;
 RX SEQUENCE FROM N.A.
 RC STRAIN=R22;
 RC MEDLINE=96171523; PubMed=8598061;
 RA Walethner J.A., Xu F., Brown K.M., Brown S.H., Gollightly E.J.,
 RA Halkier T., Kauppinen S., Pederson A., Schneider P.;
 RT 'The identification and characterization of four laccases from the
 RT plant pathogenic fungus Rhizoctonia solani.';
 RL Plant. Genet. 29:395-403(1996).
 CC -1- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
 CC PRODUCTS (PROBABLE).
 CC -1- CATALYTIC ACTIVITY: 4 BENZENEDIOL + O(2) = 4 BENZOSEMIQUINONE +
 CC 2 H(2)O.
 CC -1- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
 CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
 CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- TISSUE SPECIFICITY: IN MICELIA, AT A LOWER LEVEL THAN LCC4.
 CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
 CC -1- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z54276; IASP.
 DR HSSP: P37064; IASP.
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR002355; MultiCu oxidase2.
 DR Pfam: PF00394; Cu-oxidase; 2.
 DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 1.
 DR PROSITE: PS00080; MULTICOPPER_OXIDASE2; FALSE_NEG.
 KW Oxidoreductase; Signal; Copper; Metal-binding; Glycoprotein; Repeat;
 KW Lignin degradation; Multigene family.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 599 LACCASE 2.
 FT DOMAIN 21 145 PLASTOCYANIN-LIKE 1.
 FT DOMAIN 157 307 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 450 567 PLASTOCYANIN-LIKE 3.
 FT METAL 82 82 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 84 84 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 127 127 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 129 129 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 497 497 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 500 500 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 502 502 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 549 549 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 550 550 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 551 551 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 551 551 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 551 551 COPPER (TYPE 1) (BY SIMILARITY).
 FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 443 443 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 599 AA; 66540 MW; F01DD5C52D0CB91 CRC64;

Query Match 7.9%; Score 89; DB 1; Length 599;
 Best Local Similarity 22.7%; Pred. No. 17;
 Matches 60; Conservative 28; Mismatches 104; Indels 72; Gaps 15;

QY 3 GSHHHHGGSDKVAATITSGKSTNTVHKSSEAGTSVFYKTDMLPEDT--THVWF 60
 DB 122 GTYWHSHLSQYV---DGLRGPVYVDPHDYRN---YDVDD--ERTVFTLADWY- 170
 QY 61 NINNEKSYVSKDI--TIKDQ--IQGGQQLDLSTLINVTGTHSNYS----- 103
 DB 171 HTPESEALIAHVDVAKTIPDSGTINGKGYDPASANTNT-TLENLYLKVKKRRRLRI 229
 QY 104 -GGSATIDFEKAPFGSKITVDN-----TK-----NTIDVTIPGYSY-----SY--- 140
 DB 230 INNSALASFFRGVGHKCTIIEADVLTIRPIEVDADILAGQYRSCILKADDPDPSWIN 289
 QY 141 ---NSFSINKTKITNDQKE-----FYVNS--QAMVQEHKEEVNGKSEFNH 182
 DB 290 APTTNVLTNTVQALLVYEDDKRPTHYWKPKFLWTKISNETIYQWQHKGSHGKGGHHH 349
 QY 183 TVNINANAGIECTVAGELVKLO 206
 DB 350 KVALIGVSGLSRVKSRASDLK 373

Search completed: January 30, 2002, 00:11:31
 Job time: 544 sec

THIS PAGE BLANK (USPTO)


```

Db      226 QKARFTSFVRYTSTTTEGQHATFENSYDINYLNNDAATNEKNTSOV-----KNVF 280
Qy      193 IEGTVKGEKLV 203
Db      281 VEGEASGNQNV 291

RESULT  2
Q9F866  PRELIMINARY; PRT; 721 AA.
ID      Q9F866
AC      Q9F866;
DT      01-MAR-2001 (TREMblrel. 16, Created)
DT      01-MAR-2001 (TREMblrel. 16, last sequence update)
DT      01-JUN-2001 (TREMblrel. 17, last annotation update)
DE      COLLAGEN ADHESIN PRECURSOR.
GN      ACE.
OS      Enterococcus faecalis (Streptococcus faecalis).
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC      Enterococcus.
OX      NCBI_TaxID=1351;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-ENDS/TX0045;
RX      MEDLINE=20407335; PubMed=10948146;
RA      Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT      "Diversity of ace, a gene encoding a microbial surface component
RT      recognizing adhesive matrix molecules, from different strains of
RT      enterococcus faecalis and evidence for production of ace during human
RT      infections.";
RL      Infect. Immun. 68:5210-5217(2000).
DR      EMBL: AF260873; AAG23932.1; -.
DR      InterPro: IPR001899; Gram_pos_anchor.
DR      Pfam: PF00746; Gram_pos_anchor; 1.
DR      PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW      Signal.
FT      SIGNAL.
SQ      SEQUENCE 721 AA; 79047 MW; F09EDBE49E5DF866 CRC64;

Query Match      16.1%; Score 181; DB 2; Length 721;
Best Local Similarity 28.8%; Pred. No. 1e-05;
Matches 55; Conservative 32; Mismatches 88; Indels 16; Gaps 6;

Qy      18 TITSGNKSTNVYHKSEAG---TSSVFYKTKGMDLPEDTHVRFMLINNEKSYVSKDIT 74
AC      Q9F867;
DT      01-MAR-2001 (TREMblrel. 16, Created)
DT      01-MAR-2001 (TREMblrel. 16, last sequence update)
DT      01-JUN-2001 (TREMblrel. 17, last annotation update)
DE      COLLAGEN ADHESIN PRECURSOR.
GN      ACE.
OS      Enterococcus faecalis (Streptococcus faecalis).
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC      Enterococcus.
OX      NCBI_TaxID=1351;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-OG1RF;
RX      MEDLINE=20407335; PubMed=10948146;
RA      Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT      "Diversity of ace, a gene encoding a microbial surface component
RT      recognizing adhesive matrix molecules, from different strains of
RT      enterococcus faecalis and evidence for production of ace during human
RT      infections.";
RL      Infect. Immun. 68:5210-5217(2000).
DR      EMBL: AF260872; AAG23931.1; -.
DR      InterPro: IPR001899; Gram_pos_anchor.
DR      Pfam: PF00746; Gram_pos_anchor; 1.
DR      PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW      Signal.
FT      SIGNAL.
SQ      SEQUENCE 721 AA; 78842 MW; 1CA8AID98CE69392 CRC64;

Query Match      16.0%; Score 179; DB 2; Length 627;
Best Local Similarity 28.3%; Pred. No. 1.2e-05;
Matches 54; Conservative 33; Mismatches 88; Indels 16; Gaps 6;

Qy      18 TITSGNKSTNVYHKSEAG---TSSVFYKTKGMDLPEDTHVRFMLINNEKSYVSKDIT 74
AC      Q9F867;
DT      01-MAR-2001 (TREMblrel. 16, Created)
DT      01-MAR-2001 (TREMblrel. 16, last sequence update)
DT      01-JUN-2001 (TREMblrel. 17, last annotation update)
DE      COLLAGEN ADHESIN PRECURSOR.
GN      ACE.
OS      Enterococcus faecalis (Streptococcus faecalis).
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC      Enterococcus.
OX      NCBI_TaxID=1351;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-OG1RF;
RX      MEDLINE=20407335; PubMed=10948146;
RA      Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT      "Diversity of ace, a gene encoding a microbial surface component
RT      recognizing adhesive matrix molecules, from different strains of
RT      enterococcus faecalis and evidence for production of ace during human
RT      infections.";
RL      Infect. Immun. 68:5210-5217(2000).
DR      EMBL: AF260872; AAG23931.1; -.
DR      InterPro: IPR001899; Gram_pos_anchor.
DR      Pfam: PF00746; Gram_pos_anchor; 1.
DR      PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW      Signal.
FT      SIGNAL.
SQ      SEQUENCE 627 AA; 68848 MW; 78C5345DFD99F528 CRC64;

Qy      135 QGYGSYNSFSINKTKIT--NEOQKEFVNNSQAWYOEHGKEEYNGKSFNHTVINANAG 192
AC      Q9F867;
DT      01-MAR-2001 (TREMblrel. 16, Created)
DT      01-MAR-2001 (TREMblrel. 16, last sequence update)
DT      01-JUN-2001 (TREMblrel. 17, last annotation update)
DE      COLLAGEN ADHESIN PRECURSOR.
GN      ACE.
OS      Enterococcus faecalis (Streptococcus faecalis).
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC      Enterococcus.
OX      NCBI_TaxID=1351;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-OG1RF;
RX      MEDLINE=20407335; PubMed=10948146;
RA      Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT      "Diversity of ace, a gene encoding a microbial surface component
RT      recognizing adhesive matrix molecules, from different strains of
RT      enterococcus faecalis and evidence for production of ace during human
RT      infections.";
RL      Infect. Immun. 68:5210-5217(2000).
DR      EMBL: AF260873; AAG23932.1; -.
DR      InterPro: IPR001899; Gram_pos_anchor.
DR      Pfam: PF00746; Gram_pos_anchor; 1.
DR      PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW      Signal.
FT      SIGNAL.
SQ      SEQUENCE 721 AA; 79047 MW; F09EDBE49E5DF866 CRC64;

Qy      266 RDKARFTSFVRYTSTTTEGQHATFENSYDINYLNNDAATNEKNTSOV-----KNVF 320
Db      321 VEGEASGNQNV 331

```

```

OX      NCBI_TaxID=1351;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-MD9/TX0249;
RX      MEDLINE=20407335; PubMed=10948146;
RA      Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT      "Diversity of ace, a gene encoding a microbial surface component
RT      recognizing adhesive matrix molecules, from different strains of
RT      enterococcus faecalis and evidence for production of ace during human
RT      infections.";
RL      Infect. Immun. 68:5210-5217(2000).
DR      EMBL: AF260877; AAG23935.1; -.
DR      InterPro: IPR001899; Gram_pos_anchor.
DR      Pfam: PF00746; Gram_pos_anchor; 1.
DR      PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW      Signal.
FT      SIGNAL.
SQ      SEQUENCE 627 AA; 68848 MW; 78C5345DFD99F528 CRC64;

Query Match      16.0%; Score 179; DB 2; Length 627;
Best Local Similarity 28.3%; Pred. No. 1.2e-05;
Matches 54; Conservative 33; Mismatches 88; Indels 16; Gaps 6;

Qy      18 TITSGNKSTNVYHKSEAG---TSSVFYKTKGMDLPEDTHVRFMLINNEKSYVSKDIT 74
AC      Q9F867;
DT      01-MAR-2001 (TREMblrel. 16, Created)
DT      01-MAR-2001 (TREMblrel. 16, last sequence update)
DT      01-JUN-2001 (TREMblrel. 17, last annotation update)
DE      COLLAGEN ADHESIN PRECURSOR.
GN      ACE.
OS      Enterococcus faecalis (Streptococcus faecalis).
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC      Enterococcus.
OX      NCBI_TaxID=1351;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-OG1RF;
RX      MEDLINE=20407335; PubMed=10948146;
RA      Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT      "Diversity of ace, a gene encoding a microbial surface component
RT      recognizing adhesive matrix molecules, from different strains of
RT      enterococcus faecalis and evidence for production of ace during human
RT      infections.";
RL      Infect. Immun. 68:5210-5217(2000).
DR      EMBL: AF260872; AAG23931.1; -.
DR      InterPro: IPR001899; Gram_pos_anchor.
DR      Pfam: PF00746; Gram_pos_anchor; 1.
DR      PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW      Signal.
FT      SIGNAL.
SQ      SEQUENCE 627 AA; 68842 MW; 1CA8AID98CE69392 CRC64;

Qy      135 QGYGSYNSFSINKTKIT--NEOQKEFVNNSQAWYOEHGKEEYNGKSFNHTVINANAG 192
AC      Q9F867;
DT      01-MAR-2001 (TREMblrel. 16, Created)
DT      01-MAR-2001 (TREMblrel. 16, last sequence update)
DT      01-JUN-2001 (TREMblrel. 17, last annotation update)
DE      COLLAGEN ADHESIN PRECURSOR.
GN      ACE.
OS      Enterococcus faecalis (Streptococcus faecalis).
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC      Enterococcus.
OX      NCBI_TaxID=1351;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-OG1RF;
RX      MEDLINE=20407335; PubMed=10948146;
RA      Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT      "Diversity of ace, a gene encoding a microbial surface component
RT      recognizing adhesive matrix molecules, from different strains of
RT      enterococcus faecalis and evidence for production of ace during human
RT      infections.";
RL      Infect. Immun. 68:5210-5217(2000).
DR      EMBL: AF260873; AAG23932.1; -.
DR      InterPro: IPR001899; Gram_pos_anchor.
DR      Pfam: PF00746; Gram_pos_anchor; 1.
DR      PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW      Signal.
FT      SIGNAL.
SQ      SEQUENCE 721 AA; 79047 MW; F09EDBE49E5DF866 CRC64;

Qy      266 RDKARFTSFVRYTSTTTEGQHATFENSYDINYLNNDAATNEKNTSOV-----KNVF 320
Db      321 VEGEASGNQNV 331

```

Query Match	16.0%;	Score 179;	DB 2;	Length 721;
Best Local Similarity	28.3%;	Pred. No. 1.5e-05;		
Matches 54;	Conservative 33;	Mismatches 86;	Indels 16;	Gaps 6;
Qy 18	TITSGNKSTNTVHKSEAG---TSSVFYKTKGDMIPEDTTHRWFLININNEKSYSKDIT 74			
Dp 152	TATAATGRLTIGSVTMTERTGQIERDYPFFYKGLDAGE-SNGVRWFLNVLNNSDVTEDIS 210			
Qy 75	IKDQIQGGQQLDLSTLININVTGTHSNITYSGSALITDFEKAFFGSKITYDNTKNTIDVTIP 134			
Dp 211	IADRGSSQQLNKESEFTDIYNDKETKY---ISLAEFDEOGYGKIDFV--TNDENLREY 265			
Qy 135	OGYGSYNSFSINYKTKIT--NEQKEAFNNSQAWQOEKGKEVNGKSFNHTYHNLINANG 192			
Dp 266	RNKAFNFTFYIYATISTTEAGCHQATFENSIDINYILNNQDATNEKNTSQV-----KNVF 320			
Qy 193	IEGTVGELKY 203			
Dp 321	VEGASGNQNV 331			
RESULT 5				
Q9F857	PRELIMINARY;	PRT;	319 AA.	
AC Q9F857;				
DT 01-MAR-2001 (Tremblrel. 16, Created)				
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)				
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)				
DE COLLAEN ADHESIN PRECURSOR (FRAGMENT).				
GN ACE.				
OS Enterococcus faecalis (Streptococcus faecalis).				
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;				
OC Enterococcus.				
OX NCBI_TaxID=1351;				
RN [1]				
RP SEQUENCE FROM N.A.				
RC STRAIN-PARKER/TX2619;				
RX MEDLINE=20407335; PubMed=10948146;				
RA Nallapareddy S.R., Singh K.V., Dub R.W., Weinstein G.M., Murray B.E.;				
RT "Diversity of ace, a gene encoding a microbial surface component				
RT recognizing adhesive matrix molecules, from different strains of				
RT enterococcus faecalis and evidence for production of ace during human				
RT infections.";				
RL Infect. Immun. 68:5210-5217(2000).				
DR EMBL: AF260894; AAC23952.1; -.				
FT NON_TER 1				
FT 319 319				
SO SEQUENCE 319 AA; 36001 MW; 2054C2B1IEC5920 CRC64;				

	Query Match	15.9%	Score 178;	DB 2;	Length 319;	
	Best Local Similarity	28.3%	Pred. No.	6.5e-06;		
	Matches	54;	Conservative	33;	Mismatches	86; Indels 16; Gaps 6
OY	18 TITSGNKSTNVVHKSEAG---TSSVFYYKTDMLPEDTHVRWFLINNEKSYSKDIT	74				
Dd	112 TATARQRLTIENVTEIRGDIENDRPPFPKYVDIAGE-SNQWRWLNLNLSKDYTEIDS	170				
OY	75 IKDOIQGGQQDLDSLTLNIINVGTSHSNYSGSAITDFEKAFFPSKITVDNTKNTIDVTIP	134				
Dd	171 IADRGSGQLNKESFTFDIVNDKETRY---ISLAEFEGQGKGKIDEV--TINDRFILARY	225				
OY	135 QGYGSYNFSFNKKTKIT--NEOKEFYNNQSQAAMYOEHGKEEVNGKSFHHVHTVNINANAG	192				
Dd	226 RDKAFRTSFIYRVYTSLTEACGHQATFEHNSYDINYLNMQDATNEKNTSQV-----KNVF	280				
OY	193 IEGITVKGEIKLV 203					
Dd	281 VEGEASGNQNV 291					

ID	Q9F856	PRELIMINARY;	PRT;	319 AA.
AC	Q9F856.			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	COLLAGEN ADHESIN PRECURSOR (FRAGMENT).			
GN	ACE.			
OS	Enterococcus faecalis (Streptococcus faecalis).			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;			
OC	Enterococcus.			
OX	NCBI_TaxID=1351;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-B-343/7X2783;			
RC	MEDLINE=20407335; PubMed=10948146;			
RA	Nallareddy S.R., Singh K.V., Duh R.W., Weinstein G.M., Murray B.E.;			
RT	"Diversity of ace, a gene encoding a microbial surface component			
RT	recognizing adhesive matrix molecules, from different strains of			
RT	enterococcus faecalis and evidence for production of ace during human			
RT	infections.";			
RL	Infect. Immun. 68:5210-5217(2000).			
DR	EMBL; AF260896; AAC33954.1; -.			
FT	NON_TER	1		
FT	NON_TER	1		
FT	NON_TER	319		
FT	NON_TER	319		
SO	SEQUENCE	319 AA;	36073 MW;	94562E9DF88988A CRC64;
Query Match	15.9%;	Score 178;	DB 2;	Length 319;
Best Local Similarity	28.3%;	Pred. No. 6.5e-06;		
Matches	54;	Conservative	33;	Mismatches 86;
				Indels 16;
				Gaps 6

```

Query Match Summary      15.9%; Score 178; DB 2; Length 319;
Best Local Similarity    28.3%; Pred. No. 6, 5e-06;
Matches    54; Conservative    33; Mismatches    88; Indels    16; Gaps    6

QY      18 TTSGNKSTNVVHKSEAG--TSSVFYKTKDMLPEDTHVHFWFLINNEKSYSKDIT 74
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      112 TATATGRLTIEGVNTETNGQIERDYPFFYKVGDLAGE--SNQYRFLNVLNKSVDYEDIS 170
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      75 TKDQIQGGQGLDLSTLNLNVGTGTHSNYSGGSALIDFEKAFGSKITVYNTKNTIDVYTP 134
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      171 IADRGSGGQLKESFEPDIVDKETKY---ISLAEFGQGYGKIDFV--TDNDFMLREY 225
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      135 OGYSVSNISNYKTKRT--NEOQKEFVNSQAGAOEHEKFEVNGKSPFHHYTHINANNG 192
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      226 RKAAPFTSTVYRTSTITTDAGQHQTFFNSYDINTIOLNODATNEKNTSQV-----KNVF 280
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      193 IEGTVKGELKV 203
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      281 VEGEASGNQNV 291

RESULT      7
ID          09EU070      PRELIMINARY;      PRT;      319 AA.
AC          09EU070;
DT          01-MAR-2001 (TREMBLrel. 16, Created)
DT          01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT          01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE          COLLAGEN ADHESIN PRECURSOR (FRAGMENT).
GN          ACE.
OS          Enterococcus faecalis (Streptococcus faecalis).
OC          Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC          Enterococcus.
OX          NCBI_TaxID=1351;
RN          [1]
RP          SEQUENCE FROM N.A.
RC          STRAIN-VARIOUS STRAINS;
RX          MEDLINE=20407335; PubMed=10948146;
RA          Nallaperreddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT          "Diversity of ace, a gene encoding a microbial surface component
RT          recognizing adhesive matrix molecules, from different strains of
RT          enterococcus faecalis and evidence for production of ace during human
RT          infections."
RL          Infect. Immun. 68:5210-5217(2000).
DR          EMBL, AF260895; AAG23953.1; -.
DR          EMBL, AF260880; AAG23938.1; -.

```

DR EMBL: AF260881; AAG23939.1; -
 DR EMBL: AF260882; AAG23940.1; -
 DR EMBL: AF260883; AAG23941.1; -
 DR EMBL: AF260888; AAG23946.1; -
 DR EMBL: AF260890; AAG23948.1; -
 DR EMBL: AF260893; AAG23951.1; -
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 319 AA; 36031 MW; 2PCF56B1IAC90D20 CRC64;

Query Match 15.9%; Score 178; DB 2; Length 319;
 Best Local Similarity 28.3%; Pred. No. 6.5e-06;
 Matches 54; Conservative 33; Mismatches 88; Indels 16; Gaps 6;

OY 18 TTSGNKSNTVYHKSEAG---TSSVFYKTKGMDLPEDTTHVFMFLINNEKSYVSKDIT 74
 DB 112 TATATGRLTEGVTNTEGTQIENDYFFFKVGLAGE-SNOVWFMFLINLNSKSDYEDIS 170
 OY 75 IKDQIOGGQQLDLSTLINVTGTHSNYSGQSAITDFEKAFFPSKITVDNKTITDVTIP 134
 DB 171 IADROSGGQQLNKSEFTFIVNDKETRY--ISLAEFEQGGYCKIDFV--TDNDFMLRFY 225
 OY 135 OGGSYNSFSINKTKIT--NEQKFEVNNNSQAWYOEHEKKEEYNGKSFNHTYHNANAG 192
 DB 226 ROKARFTSFIVRTSTITTEAGQHQAATFENSYDINTYOLNODATNEKNTSOV-----KNVF 280
 OY 193 IEGTVKGLKV 203
 DB 281 VEGEASGNONV 291

RESULT 8
 OYXB07 PRELIMINARY; PRT; 458 AA.
 AC OYXB07:
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE COLLAGEN ADHESIN ACE (FRAGMENT).
 GN ACE.
 OS Enterococcus faecalis (Streptococcus faecalis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
 OC Enterococcus.
 OX NCBI_TaxID=1351;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CG110;
 RA Rich R.L., Kreikemeier B., Owens R.T., Labrenz S., Narayana S.V.L.,
 RA Weinstock G.W., Murray B.E., Hook M.,
 RT "Ace: a collagen-binding MSCRAMM from Enterococcus faecalis.";
 RL J. Biol. Chem. 0:0-0(1999).
 DR EMBL: AF159247; AAD43342.1; -
 DR HSSP: O53654; IAMX.
 DR COLLAGEN.
 KM
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 458 AA; 51085 MW; B674692BD29A0A24 CRC64;

Query Match 15.9%; Score 178; DB 2; Length 458;
 Best Local Similarity 28.3%; Pred. No. 1e-05;
 Matches 54; Conservative 33; Mismatches 88; Indels 16; Gaps 6;

OY 18 TTSGNKSNTVYHKSEAG---TSSVFYKTKGMDLPEDTTHVFMFLINNEKSYVSKDIT 74
 DB 121 TATATGRLTEGVTNTEGTQIENDYFFFKVGLAGE-SNOVWFMFLINLNSKSDYEDIS 179
 OY 75 IKDQIOGGQQLDLSTLINVTGTHSNYSGQSAITDFEKAFFPSKITVDNKTITDVTIP 134
 DB 180 IADROSGGQQLNKSEFTFIVNDKETRY--ISLAEFEQGGYCKIDFV--TDNDFMLRFY 234
 OY 135 OGGSYNSFSINKTKIT--NEQKFEVNNNSQAWYOEHEKKEEYNGKSFNHTYHNANAG 192

DB 235 ROKARFTSFIVRTSTITTEAGQHQAATFENSYDINTYOLNODATNEKNTSOV-----KNVF 289
 OY 193 IEGTVKGLKV 203
 DB 290 VEGEASGNONV 300

RESULT 9
 OYF862 PRELIMINARY; PRT; 580 AA.
 AC OYF862:
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE COLLAGEN ADHESIN PRECURSOR.
 GN ACE.
 OS Enterococcus faecalis (Streptococcus faecalis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
 OC Enterococcus.
 OX NCBI_TaxID=1351;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LBJ-1/TK0020;
 RX MEDLINE=20407335; PubMed=10948146;
 RX Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.W., Murray B.E.;
 RT "Diversity of ace, a gene encoding a microbial surface component
 RT recognizing adhesive matrix molecules, from different strains of
 RT enterococcus faecalis and evidence for production of ace during human
 RT infections.";
 RL Infect. Immun. 68:5210-5217(2000).
 RL EMBL: AF260878; AAG23936.1; -
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 KW Signal.
 FT SIGNAL. 1
 FT SIGNAL. 31
 FT SIGNAL. POTENTIAL.
 SQ SEQUENCE 580 AA; 63948 MW; 7CD61EAD3FDA0993 CRC64;

Query Match 15.9%; Score 178; DB 2; Length 580;
 Best Local Similarity 28.3%; Pred. No. 1.3e-05;
 Matches 54; Conservative 33; Mismatches 88; Indels 16; Gaps 6;

OY 18 TTSGNKSNTVYHKSEAG---TSSVFYKTKGMDLPEDTTHVFMFLINNEKSYVSKDIT 74
 DB 152 TATATGRLTEGVTNTEGTQIENDYFFFKVGLAGE-SNOVWFMFLINLNSKSDYEDIS 210
 OY 75 IKDQIOGGQQLDLSTLINVTGTHSNYSGQSAITDFEKAFFPSKITVDNKTITDVTIP 134
 DB 211 IADROSGGQQLNKSEFTFIVNDKETRY--ISLAEFEQGGYCKIDFV--TDNDFMLRFY 265
 OY 135 OGGSYNSFSINKTKIT--NEQKFEVNNNSQAWYOEHEKKEEYNGKSFNHTYHNANAG 192
 DB 266 ROKARFTSFIVRTSTITTEAGQHQAATFENSYDINTYOLNODATNEKNTSOV-----KNVF 320
 OY 193 IEGTVKGLKV 203
 DB 321 VEGEASGNONV 331

RESULT 10
 OYF861 PRELIMINARY; PRT; 580 AA.
 AC OYF861:
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE COLLAGEN ADHESIN PRECURSOR.
 GN ACE.
 OS Enterococcus faecalis (Streptococcus faecalis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
 OC Enterococcus.


```

RESULT 13
ID 09F858 PRELIMINARY; PRT: 319 AA.
AC 09F858:
DB 01-MAR-2001 (TREMBlrel. 16, Created)
DB 01-MAR-2001 (TREMBlrel. 16, last sequence update)
DB 01-MAR-2001 (TREMBlrel. 16, last annotation update)
DE COLLAGEN ADHESIN PRECURSOR (FRAGMENT).
GN ACE.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BE83/TX0855;
RX MEDLINE=20407335; PubMed=10948146;
RA Nallapereddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT "Diversity of ace, a gene encoding a microbial surface component
RT recognizing adhesive matrix molecules, from different strains of
RT enterococcus faecalis and evidence for production of ace during human
RT infections.";
RL Infect. Immun. 68:5210-5217(2000).
DR EMBL; AF260891; AAG23949.1; -.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 319 AA; 36096 MW; 7716569F98CA593C CRC64;

```

```

Query Match 15.8%; Score 177.5; DB 2; Length 319;
Best Local Similarity 25.0%; Pred. No. 7.1e-06;
Matches 53; Conservative 39; Mismatches 83; Indels 37; Gaps 7;

```

```

QY 16 VATTSGNSTVTVHKSSEAG---TSSVYFKYKTDMLPEDTTHVRFMLINNEKSYVSKDIT 53
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 93 IKTLLTNSQPNVT--ETDFGTAITATQRLTIEGNTTEMGOIEBODYFFKYGDLDGE-S 149
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 54 THVRWFLINNEKSYVSKDITTKDIOGQOGLDLSTLININVTGHSNYSYGOSAITDPER 113
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 150 NGVRWFLINNLKNSDVTEDISADRGSGQOLNKSEFTFDIVDKETKY---ISLAEFQ 206
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 114 APPGSKITVDNKTNTDVTIPQGYGSYNFSINTKKIT--NEQKEFYNNNSQATYQEHG 171
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 207 OGYGKITYE--TDNDFNLFFYRKARFTSFIVKYSTITEAGHQATFENSYDINQLNN 264
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 172 KEEVNCKSPNHTVHNINANAGIEGTVKGEIKY 203
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 265 QDATNEKNTSQV-----KNVFEVGEASGNQNV 291
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

```

```

RESULT 14
ID 09F860 PRELIMINARY; PRT: 296 AA.
AC 09F860:
DB 01-MAR-2001 (TREMBlrel. 16, Created)
DB 01-MAR-2001 (TREMBlrel. 16, last sequence update)
DB 01-MAR-2001 (TREMBlrel. 16, last annotation update)
DE COLLAGEN ADHESIN PRECURSOR (FRAGMENT).
GN ACE.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WH245/TX0635;
RX MEDLINE=20407335; PubMed=10948146;
RA Nallapereddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT "Diversity of ace, a gene encoding a microbial surface component
RT recognizing adhesive matrix molecules, from different strains of
RT enterococcus faecalis and evidence for production of ace during human
RT infections.";

```

```

RL Infect. Immun. 68:5210-5217(2000).
DR EMBL; AF260887; AAG23945.1; -.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 296 AA; 33331 MW; 8D9630648275715F CRC64;

```

```

Query Match 15.6%; Score 175; DB 2; Length 296;
Best Local Similarity 27.7%; Pred. No. 1e-05;
Matches 53; Conservative 34; Mismatches 88; Indels 16; Gaps 6;

```

```

QY 18 TTSGNKSNTVTVHKSSEAG---TSSVYFKYKTDMLPEDTTHVRFMLINNEKSYVSKDIT 74
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 89 TATATQRLTIEGVTNTEQIERDYFFKYVGDLDGE--SNQVRWFLINNLKNSDVTEDIS 147
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 75 IKDIOGQOGLDLSTLININVTGHSNYSYGOSAITDPERKAFPGSKITVDNKTNTDVTIP 134
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 148 IADRGSGQOLNKSEFTFDIVDKETKY---ISLAEFEQGYGKIDFV--TDNDFNLRY 202
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 135 OGYSYNSFSINVKTKIT--NEQKEFYNNNSQAWYOEHEKKEVNGKSPNHTVHNINANG 192
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 203 RDKARTSFIVKYSTITEAGHQATFKNSYDINQLNNDATNEKNTSQV-----KNVF 257
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 193 IEGTVKGEIKY 203
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 258 VDGESAGNQN 268
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

```

```

RESULT 15
ID 09E080 PRELIMINARY; PRT: 319 AA.
AC 09E080:
DB 01-MAR-2001 (TREMBlrel. 16, Created)
DB 01-MAR-2001 (TREMBlrel. 16, last sequence update)
DB 01-MAR-2001 (TREMBlrel. 16, last annotation update)
DE COLLAGEN ADHESIN PRECURSOR (FRAGMENT).
GN ACE.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BE88/TX0860; HG6280/TX0630, HG10528/TX0631, AND HG9829/TX0633;
RX MEDLINE=20407335; PubMed=10948146;
RA Nallapereddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E.;
RT "Diversity of ace, a gene encoding a microbial surface component
RT recognizing adhesive matrix molecules, from different strains of
RT enterococcus faecalis and evidence for production of ace during human
RT infections.";
RL Infect. Immun. 68:5210-5217(2000).
DR EMBL; AF260892; AAG23950.1; -.
DR EMBL; AF260884; AAG23942.1; -.
DR EMBL; AF260885; AAG23943.1; -.
DR EMBL; AF260886; AAG23944.1; -.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 319 AA; 35942 MW; 296BCB542350DA8A CRC64;

```

```

Query Match 15.6%; Score 175; DB 2; Length 319;
Best Local Similarity 27.7%; Pred. No. 1.1e-05;
Matches 53; Conservative 34; Mismatches 88; Indels 16; Gaps 6;

```

```

QY 18 TTSGNKSNTVTVHKSSEAG---TSSVYFKYKTDMLPEDTTHVRFMLINNEKSYVSKDIT 74
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 112 TATATQRLTIEGVTNTEQIERDYFFKYVGDLDGE--SNQVRWFLINNLKNSDVTEDIS 170
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 75 IKDIOGQOGLDLSTLININVTGHSNYSYGOSAITDPERKAFPGSKITVDNKTNTDVTIP 134
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 171 IADRGSGQOLNKSEFTFDIVDKETKY---ISLAEFEQGYGKIDFV--TDNDFNLRY 225
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 135 OGYSYNSFSINVKTKIT--NEQKEFYNNNSQAWYOEHEKKEVNGKSPNHTVHNINANG 192
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

```

Db	226	RDKARFTSFIVRTSTITEAGQHQAPEKNSYDINQOLNODATNEKNTSQV-----KNVF	280
QY	193	IEGTWKGELKV	203
Db	281	VDGEASGNCNV	291

Search completed: January 30, 2002, 00:10:29
Job time: 557 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2002, 22:10:31 ; Search time 167.16 Seconds
(without alignments)
4354.326 Million cell updates/sec

Title: US-09-813-820-3

Perfect score: 849
Sequence: 1 GACGATTAATAATGAAAAAT.....AACAGCATTAAGATACCAAG 849

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq.1101.*
1: /SID8/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID8/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID8/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID8/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID8/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID8/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID8/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID8/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID8/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID8/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID8/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID8/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID8/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID8/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID8/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID8/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID8/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID8/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID8/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID8/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID8/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID8/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	849	100.0	849	19	AAAT93437
2	849	100.0	1500	19	AAAT93438
3	790.6	93.1	4612	13	AAO24123
4	441	51.9	441	19	AAAT93436
5	170	20.0	936	22	AAFS8252
6	170	20.0	936	22	AAFS8254
7	170	20.0	936	22	AAFS8257
8	170	20.0	936	22	AAFS8259
9	170	20.0	936	22	AAFS8262
10	170	20.0	938	22	AAFS8255
11	162.4	19.1	936	22	AAFS8252

12	162.4	19.1	936	22	AAFS8254	Oligonucleotide D1
13	162.4	19.1	936	22	AAFS8257	Oligonucleotide D1
14	162.4	19.1	936	22	AAFS8259	Oligonucleotide D2
15	162.4	19.1	936	22	AAFS8262	Oligonucleotide D2
16	162.4	19.1	938	22	AAFS8255	Oligonucleotide D1
17	162.4	7.1	3869	20	AAAT1327	Enterococcus faec
18	58.8	6.9	244	22	AAFS8238	Oligonucleotide D1
19	58.6	6.9	244	22	AAFS8238	Oligonucleotide D1
20	57	6.7	1374	24	AAAC83828	Collagen binding A
21	51	6.0	6644	20	AAAT33181	Base sequence of t
22	51	6.0	7372	20	AAAT33182	Base sequence of t
23	51	6.0	7797	20	AAAT33180	Cowpox virus bsr f
24	51	6.0	7996	20	AAAT33184	Base sequence of t
25	50.8	6.0	1524	21	AAAT70107	Plasmodium falci
26	50.6	6.0	15016	20	AAAT99560	Nucleic acid seque
27	49.4	5.8	3837	21	AAAT70211	Plasmodium falci
28	49.4	5.8	6033	21	AAAT70152	Plasmodium falci
29	49	5.8	3159	22	AAAC93000	Candida albicans C
30	48.4	5.7	1677	21	AAAT70207	Plasmodium falci
31	48.4	5.7	1865	21	AAAT70096	Plasmodium falci
32	47.6	5.6	5454	21	AAAT70236	Plasmodium falci
33	47.4	5.6	19124	18	AAAT72882	Plasmodium var-7 g
34	47.4	5.6	19124	18	AAAT72882	Plasmodium var-7 g
35	47	5.5	3567	21	AAAT70117	Plasmodium falci
36	46.8	5.5	1308	20	AAAT99503	Nucleic acid seque
37	46.8	5.5	1683	21	AAAT70112	Plasmodium falci
38	46.6	5.5	4041	21	AAAT70170	Plasmodium falci
39	46.4	5.5	1998	21	AAAT70212	Plasmodium falci
40	46.2	5.4	7458	21	AAAT70106	Plasmodium falci
41	46	5.4	3095	11	AAAT03875	Sequence encoding
42	46	5.4	5940	21	AAAT70105	Plasmodium falci
43	46	5.4	7491	17	AAAT70156	Plasmodium falci
44	46	5.4	9789	17	AAAT41852	cDNA encoding plas
45	45.8	5.4	4590	22	AAAT24065	Yeast AOD9604-asso

ALIGNMENTS

RESULT 1	AAAT93437	standard: DNA; 849 BP.
ID	AAAT93437	
AC	AAAT93437	
XX		
DT	21-MAY-1998	(first entry)
XX		
DE	Collagen binding protein M31 epitope DNA.	
XX		
KW	Collagen binding protein: cna gene; sepsis; infection;	
KW	microbial surface component regionisng adhesive matrix molecule;	
KW	MSCRAMM; adhesin; vaccine; immunisation; diagnosis; therapy;	
KW	epitope M31; ss.	
XX		
OS	Staphylococcus aureus.	
XX		
EH	Key	Location/Qualifiers
FT	CDS	253..849
FT		/*tag= a
XX		
PN	WO9743314-A2.	
XX		
PD	20-NOV-1997.	
XX		
PF	14-MAY-1997;	97WO-US08210.
XX		
PR	16-MAY-1996;	96US-0017678.
XX		
PA	(UABR-) UAB RES FOUND.	
PA	(TEXA) UNIV TEXAS A & M SYSTEM.	
XX		
PI	Hook M, House-Pompeo K, Patil JM, Sthanam N, Symersky J;	
XX		

DR MPI: 1998-008801/01.
 XX P-PSDB: AAM31553.
 PT Antibody that interacts with collagen binding domain of
 PT Staphylococcal cna gene product - useful to prevent bacterial sepsis
 PT in animal infected with Staphylococcus aureus
 XX
 PS Claim 40; Page 114-115; 143pp; English.
 XX
 CC This nucleic acid comprises a portion of the Staphylococcus aureus
 CC cna gene that codes for collagen binding protein (CBP) epitope M31
 CC (see AAM31553), i.e. amino acids 61-343 of full-length CBP. Claimed
 CC 441, 849 and 1500 bp nucleic acid sequences (see AAT93436-38)
 CC respectively encode CBP epitopes M17, M31 and M55 (see AAM31552-54)
 CC that confer protection against S. aureus infection. These nucleic
 CC acid sequences can be used in the recombinant production of the CBP
 CC epitopes. The CBP protein and antigenic epitopes are contemplated
 CC for use in the treatment of pathological infections, especially to
 CC prevent bacterial adhesion to collagen. The claimed nucleic acids
 CC as well as claimed anti-CBP antibodies will also be of use in
 CC screening, diagnostic and therapeutic applications including active
 CC and passive immunisation and methods for the prevention of
 CC bacterial colonisation in an animal such as a human. These DNA
 CC segments and the peptides encoded by them are also contemplated for
 CC use in the preparation of vaccines and as carrier proteins in
 CC vaccine formulations, as well as in the formulation of compositions
 CC for the prevention of S. aureus infection.
 CC
 XX Sequence 849 BP; 348 A; 110 C; 172 G; 219 T; 0 other;
 SQ
 Query Match 100.0%; Score 849; DB 19; Length 849;
 Best Local Similarity 100.0%; Pred. No. 9.8e-177;
 Matches 849; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GACGATAAATGGAATAATACAAAATGTCATGATTAAGTGGCATGCGCGAAGC 60
 Db 1 gacgataaaatggaataaatacaaaatggtgacatgataaagtcagtcgcaaacg 60
 QY 61 GGTACAGTAAGATAGAGGTTATAGTAATAACAGTTCATTAAGTGAAGCAAG 120
 Db 61 ggtacagtaagatagaggttatagtaataaacagtcaccattacgtttaagtgacag 120
 QY 121 GTGGGTCAAGCAGTATTACACACAGCGGTGCACAAATTACATTCAATGAAGTAGAA 180
 Db 121 gtgggtcaagcagttattacaccacagcgtgcaacaattacattcaatgataagtagaa 180
 QY 181 AATTAAGTATGATCTTCGGGATTTGCAGAAATTGAAGTCAAGGAAGAATTAAACGCCAA 240
 Db 181 aatlaagtgatcttcgggatttcgagaattgaaagtaacgaagaatttaacgcaa 240
 QY 241 ACAATACTCTCAGATGACAAAGTAGCTACGATTAACATCTGGGAATTAATCAAGATGTT 300
 Db 241 acaaatactctcagatgacaaagtagctacgataacatctcgggaataataccaagatgt 300
 QY 301 ACGGTTTCATTAAGTGAAGCGGGAACAGTAGTGTGTTCTATTATTAACCGGAGATATG 360
 Db 301 acggttcataaagtgaaagcgggaacaagtagtcttctattataaacgggaagtagatg 360
 QY 361 GTACCAAGAAGATGACGACATGTACGATGTTTAAATTAATTAACAAATTAATTTAT 420
 Db 361 gtaccaagaagatgacgacatgtacgattgtttaaataataaacaagtaaatat 420
 QY 421 GTATGAAAGATATTACTATTAAAGATCAGATTCAAGTGGACAGCAGTAGATTAAAC 480
 Db 421 gtatgaaagatattactattaaagatcagattcaagtgagcagagtagattaaagc 480
 QY 481 ACATTAACATTAAATGTGACAGGTACACATACATTAATTAATGTCGACAAAGTGCAT 540
 Db 481 acatttaaacatttaagtgcaggtacacatgacattattatagtgacaaagtgcaatt 540
 QY 541 ACTGATTTTGAAGAAAGCCTTCCAGTTCTAAATTAACGTTGATTAATCGAAGAACAA 600
 Db 541 actgattttgaaagaaagccttccagttcttaaatatcactgttgataatacgaagaacaa 600

Db 541 actgattttgaaagaaagccttccagttcttaaatatcactgttgataatacgaagaacaa 600
 QY 601 ATTGATGTAAACATTTCCACAGGCTATGGGTATATATATGTTTTCAATTAACATCAAA 660
 Db 601 attgatgttaaacatttcacaggtctatgggtatataatgTTTTTCAATTAACATCAAA 660
 QY 661 ACCAAATTTACGAATGACAGCAAAAAGGTTTGTTAAATTTACACACTTGATCA 720
 Db 661 accaaatttacgaatgacagcaaaaaggtttgTTAAATTTACACACTTGATCA 720
 QY 721 GAGCATGTAAAGAAAGTGAACGGGAATCATTTATCATCTGTCACATATTAT 780
 Db 721 gagcatgttaagaaagtgaacgggaatcatTTATCATCTGTCACATATTAT 780
 QY 781 GCTAATGCCGATTTGTAAGTACTGTAAAGGTGAATTAAGTTTAAACAGATAAA 840
 Db 781 gctaatgccgatttgaagtactgttaaaagtgtaatttaaaagtttaaacaggtataa 840
 QY 841 GATACCAAG 849
 Db 841 gataccaag 849
 RESULT 2
 AAT93438
 ID AAT93438 standard; DNA; 1500 BP.
 XX
 AC AAT93438;
 XX
 DT 21-MAY-1998 (first entry)
 XX
 DE Collagen binding protein M55 epitope DNA.
 XX
 DE Collagen binding protein; cna gene; sepsis; infection;
 KW microbial surface component regionising adhesive matrix molecule;
 KW MSCRAMM; adhesin; vaccine; immunisation; diagnosis; therapy;
 KW epitope M55; ss.
 XX
 OS Staphylococcus aureus.
 XX
 PN MO9743314-A2.
 XX
 PD 20-NOV-1997.
 XX
 PF 14-MAY-1997; 97WO-US08210.
 XX
 PR 16-MAY-1996; 96US-0017678.
 XX
 PA (UABR-) UAB RES FOUND.
 PA (TEXA) UNTV TEXAS A & M SYSTEM.
 XX
 PI Hook M, House-Pompeo K, Patil JM, Sthanam N, Symersky J;
 XX
 DR MPI: 1998-008801/01.
 DR P-PSDB: AAM31554.
 XX
 PT Antibody that interacts with collagen binding domain of
 PT Staphylococcal cna gene product - useful to prevent bacterial sepsis
 PT in animal infected with Staphylococcus aureus
 XX
 PS Claim 40; Page 116-117; 143pp; English.
 XX
 CC This nucleic acid comprises a portion of the Staphylococcus aureus
 CC cna gene that codes for collagen binding protein (CBP) epitope M55
 CC (see AAM31554), i.e. amino acids 30-531 of full-length CBP. Claimed
 CC 441, 849 and 1500 bp nucleic acid sequences (see AAT93436-38)
 CC respectively encode CBP epitopes M17, M31 and M55 (see AAM31552-54)
 CC that confer protection against S. aureus infection. These nucleic
 CC acid sequences can be used in the recombinant production of the CBP
 CC epitopes. The CBP protein and antigenic epitopes are contemplated
 CC for use in the treatment of pathological infections, especially to
 CC prevent bacterial adhesion to collagen. The claimed nucleic acids
 CC as well as claimed anti-CBP antibodies will also be of use in

screening, diagnostic and therapeutic applications including active
 CC and passive immunisation and methods for the prevention of
 CC bacterial colonisation in an animal such as a human. These DNA
 CC segments and the peptides encoded by them are also contemplated for
 CC use in the preparation of vaccines and as carrier proteins in
 CC vaccine formulations, as well as in the formulation of compositions
 CC for the prevention of *S. aureus* infection.

XX
 Sequence 1500 BP; 628 A; 198 C; 296 G; 378 T; 0 other;

Query Match 100.0%; Score 849; DB 19; Length 1500;

Best Local Similarity 100.0%; Pred. No. 1,1e-176;
 Matches 849; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACGATAAATAATGAAAAATACAAAATGTCATGATTAAGTCGATGCCGACACAGC 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 94 gacgataaataatgaaaaatacacaatggtgacatgattaaagtgcgtgcgcacaagc 153
 QY 61 GGTACGATAAGATAGAGGTTTACTTAAACACGTACCAATTACTGTAAAGGTGAACAG 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 154 ggtacgataagataagaggtatagtaaaacagtcacatcgtltaaggtgacagc 213
 QY 121 GTGGGTCAAGCAGTTTATACACGACGCGTGCACAAATTACATTCATGATAAGTAAAGAA 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 214 gtgggtcaagcagttatatacaccagcgtgacacaattacatcattgataagtagaa 273
 QY 181 AATTAAGTGATGTTTCCGGATTTCGAGAAATTTGAATGACAGGAAGAAATTTAACGCA 240
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 274 aatthaagtgatgttccggatttcgagatttcgagaatttgaagatacagaagaatttaacgcaa 333
 QY 241 ACAAAATCTTCAGATGACAAAGTACGATACATCTGGGAATTAATCAACGAATGTT 300
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 334 acaaatcttcagatgacaaagtagctacgataacalcgtggaataataatcaacgaatglt 393
 QY 301 ACGGTTCAATAAAGTAAAGCAGGACGGAACAGTAGTGTTCCTATTATTAACCGGAGATATG 360
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 394 acggttcataaaagtgaaagcgggaacaagtagtgcattccattataaaacggagatagc 453
 QY 361 CTACCGAAGAGATGACACATGTAGCATGGTTTAAATATTAAACAAATGAAAAAGTTAT 420
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 454 ctaccgaagatatacgacacatgtaagatgtttttaataattaacaatgtaaaaaagttat 513
 QY 421 GATTGGAAGATTTACTTAAAGATCAGATTCAGATGCGAGACAGATTCAGATTTAAGC 480
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 514 gtatcgaaagatactatacctaagaagtcagatcaagtgtagacagcagttagattaaagc 573
 QY 481 ACATTAAACATTAAATGTGACAGTACACATAGCAATTATTAATAGTGACAAAGTGCAT 540
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 574 acctaaacatttaattgacaggtacacataagaattatattatagtgacaaagtgcaatt 633
 QY 541 ACTGATTTTGAATAAGCCTTTCAGGTTCTAAATAACTGTTGATTAATAGGAACACACA 600
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 634 actgatttcgaaaaagccttcacaggttcctaaataactgttgataataagaagaacaca 693
 QY 601 ATTGATGATAACAATTCCACAAGGCTATGGTCTATTAATAGTTTTCATTAATCTACAA 660
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 694 atgtagtatacaattccacaaggtctatgggtcattataatagtttccaattactaaca 753
 QY 661 ACCAAATTTACGAATGACAGCAAAAAGCTTGTAAATTAATCACAAGCTGGTATCA 720
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 754 accaaataatacgaatgaacagcaaaaagatttcgttaataatacacaagcttgcgtatcaa 813
 QY 721 GAGCATGTAAGGAGAAGTGAACGGGAATCTTTAATCATCACTGTCACAAATATTAA 780
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 814 gagcatgtaagaagaagtgaaacgggaatacatttaatactatcgtgcacaattatatt 873
 QY 781 GCTAATGCCGTTATTGAAGTACTGTAAGGTTAAAGTAAAGTTTAAACGAGATTA 840
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 874 gctaatgccggtatttgaaagtaactgtataaagtgtaataaagtttaaaacaggtataa 933
 QY 841 GATACCAAG 849
 ||||||||||

DB 934 gataccaag 942
 RESULT 3
 ID AAQ24123 standard; DNA; 4612 BP.
 AC AAQ24123;
 XX
 XX 04-DEC-1992 (first entry)
 DE p16 and cCOLR6a assembled cbp sequence.
 XX CBP; collagen binding protein; ss.
 KW
 OS Staphylococcus aureus.
 XX
 XX
 FH Key Location/Qualifiers
 FT misc_RNA 1018..4485
 FT /tag= a
 FT /note= "Claim 6; page 26-28"
 FT 1..8
 FT /tag= b
 FT /note= "part of the vector MCS"
 FT misc_RNA 9..3930
 FT /tag= c
 FT /note= "Insert in p16"
 FT 2385..4612
 FT /tag= d
 FT /label=
 FT /note= "5' end of insert in cCOLR6a"
 FT 917..924
 FT /tag= e
 FT CDS 931..4488
 FT /tag= f
 FT /label= CBP
 FT 931..1017
 FT /tag= g
 FT /label= S
 FT /note= "signal sequence"
 FT misc_RNA 1018..2532
 FT /tag= h
 FT /label= A
 FT /note= "see CC"
 FT 2533..3093
 FT /tag= i
 FT /label= B1
 FT /note= "see CC"
 FT misc_RNA 3094..3654
 FT /tag= j
 FT /label= B2
 FT /note= "see CC"
 FT 3655..4215
 FT /tag= k
 FT /label= B3
 FT /note= "see CC"
 FT misc_RNS 4216..4407
 FT /tag= l
 FT /label= W
 FT /note= "see CC"
 FT 4408..4467
 FT /tag= m
 FT /label= M
 FT /note= "see CC"
 FT 4468..4485
 FT /tag= n
 FT /note= "charged C-terminal"
 PF 22-OCT-1991; 91WO-SE00207.

CC screening, diagnostic and therapeutic applications including active
CC and passive immunisation and methods for the prevention of
CC bacterial colonisation in an animal such as a human. These DNA
CC segments and the peptides encoded by them are also contemplated for
CC use in the preparation of vaccines and as carrier proteins in
CC vaccine formulations, as well as in the formulation of compositions
CC for the prevention of *S. aureus* infection.

XX
SQ Sequence 441 BP; 180 A; 60 C; 78 G; 123 T; 0 other;

Query Match 51.9%; Score 441; DB 19; Length 441;
Best Local Similarity 100.0%; Pred. No. 1.2e-87;

Matches 441; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 ATATACATCTGGGAATTAATCAAGCAATGTTACGTTATATAAGTGAAGCGGAACAACT 330
DB 1 ataccatctgggaataatacaagaaatgtaacggttcataaagtgaagcgggaacaagt 60

QY 331 AGTGTCTTATATATAAAGGAGATATGCTACCAAGATACGACACATGTACGATGG 390
DB 61 aggtttctcatcatalaacaagcgagatalatgtaccagaagatacagacacatgtaagatg 120

QY 391 TTTTAAATATTAACAATGAAAAAGTTATGTATCGAAAGATATTACTATTAAGAGATCAG 450
DB 121 tttaataataatacaatgaaaaagttatgtatcgaaagatatataataaagatcag 180

QY 451 ATTCAAGGTGACAGCACTTATTAAGCAATTAACATTTAATGTACAGGTACACAT 510
DB 181 attcaaggtgacagcaggttagatttaagcacatlaaactaatgacaggtacacat 240

QY 511 AGCAATATATATAGTGCAAGAGCAATTAAGTATTTAAAAAGCCTTCCAGGTCT 570
DB 241 agcaatataatagtcggaacagtcgaatcagatcttgaanaagccttcagagttct 300

QY 571 AAATAACTGTGTATATACGAAGAACAATTTGATGTAACAATTCACAAGGCTATGG 630
DB 301 aaataactgttgataataacgaagaacaacatitgatacaatcccaagcgtatgg 360

QY 631 TCATATATATGTTTTCATTAATTAACATAAACCATAATTCGAATGAAGCAAAAAAGAG 690
DB 361 tcataataatgatttcaactaactacaaaacaaatcagaaatgacagcaaaaagag 420

QY 691 TTGTTAATATTCACAAGCT 711
DB 421 ttgtaataataatcacaagct 441

RESULT 5
AAF58252/c
ID AAF58252 standard; DNA: 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KM Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000MO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.

XX
PI Unk RM;

XX
DR WPI; 2001-159728/16.

XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface

PS Example 6; Page 127; 159pp; English.

XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.

XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 20.0%; Score 170; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 2.1e-28;

Matches 8; Conservative 524; Mismatches 254; Indels 0; Gaps 0;

QY 59 GCGGTACGTAATAGAGGCTTATAGTAATAACAGTACCTTAAGTGAAC 118
DB 791 GCGGCTACGTAATAGAGGCTTATAGTAATAACAGTACCTTAAGTGAAC 732

QY 119 AGTGGGTCAAGCACTTATTAACACAGGTCACAACTTCAATGATTAAGTAG 178
DB 731 AGTGGGTCAAGCACTTATTAACACAGGTCACAACTTCAATGATTAAGTAG 672

QY 179 AAAAATTAAGTATGTTTCGGGATTCGCAATTTGAAGTGAAGAAATTAACGC 238
DB 671 AAAAATTAAGTATGTTTCGGGATTCGCAATTTGAAGTGAAGAAATTAACGC 612

QY 239 AAAAATTAAGTATGTTTCGGGATTCGCAATTTGAAGTGAAGAAATTAACGC 298
DB 611 AAAAATTAAGTATGTTTCGGGATTCGCAATTTGAAGTGAAGAAATTAACGC 552

QY 299 TTACGTTCAATTAAGTGAAGCGGACCAAGTATGTTTCTATTAATTAACGAGATTA 358
DB 551 TTACGTTCAATTAAGTGAAGCGGACCAAGTATGTTTCTATTAATTAACGAGATTA 492

QY 359 TGCTACGAGAAATACGACATGATGATGTTTAAATATTAACATGAAGAAAGTT 418
DB 491 TGCTACGAGAAATACGACATGATGATGTTTAAATATTAACATGAAGAAAGTT 432

QY 419 ATGTATCGAAAGATATTACTATTAAGATCAGATTCAGGACAGCAGTTAGATTAA 478
DB 431 ATGTATCGAAAGATATTACTATTAAGATCAGATTCAGGACAGCAGTTAGATTAA 372

QY 479 GCACATTAACATTAATGTGACAGTACACATACCAATTAATAGTGACAAAGTGCAA 538
DB 371 GCACATTAACATTAATGTGACAGTACACATACCAATTAATAGTGACAAAGTGCAA 312

QY 539 TTACTGATTTGAAAAGCCTTCCAGGTTCTAAATTAAGTGAATTAAGCAAGACA 598
DB 311 TTACTGATTTGAAAAGCCTTCCAGGTTCTAAATTAAGTGAATTAAGCAAGACA 252

QY 599 CAATTGATGTAACATTCACACAGGCTATGGGTCATTAATAGTTTCAATTACTACA 658
DB 251 CAATTGATGTAACATTCACACAGGCTATGGGTCATTAATAGTTTCAATTACTACA 192

QY 659 AAACCAAAATTAAGCAATGAACAGCAAAAAGGTTGTTAATTAATTCACAACTGGTATC 718
DB 191 AAACCAAAATTAAGCAATGAACAGCAAAAAGGTTGTTAATTAATTCACAACTGGTATC 132

QY 719 AAGAGCATGTAAGCAAGAGTGAACGGAATCATTTAATCATAGTGCACATATTA 778
DB 131 AAGAGCATGTAAGCAAGAGTGAACGGAATCATTTAATCATAGTGCACATATTA 72

```

QY 779 ATGCTATGCCGTATGTGAAGTACTGTATAAAGTGCAATTTAAAGTTTAAACAGAGATA 838
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 71 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 12
QY 839 AAGATA 844
  : : : : :
Db 11 WWWWWW 6

RESULT 6
AAF58254/c
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KM Electron-transfer group; ETW; mismatch; genotyping;
KM gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
PI
PI -Umek RM;
XX
WPI; 2001-159728/16.
XX
DR
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
XX
PS Example 6; Page 127; 159pp; English.
XX
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETW) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
CC
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 20.0%; Score 170; DB 22; Length 936;
Best Local Similarity 1.0%; Pired. No. 2, 1e-28;
Matches 8; Conservative 524; Mismatches 254; Indels 0; Gaps 0;

QY 59 GCGGTACAGTAAAGATGAGGGTATTAGTAAACACTACCATTAAGTGAAC 118
  | | : : : : : : : : : : : : : : : : : : : : : : : :
Db 791 GGGCGMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 732
QY 119 AGGTGGGTCAAGCAGTTATTACACGACGTCACAATTCATTCATGAATAGTAG 178
  : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 731 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 672
QY 179 AAAAATTAAGTATGTTGGGATTTGACGAATTTGAAGTACAAGAGAAATTTAAGC 238
  : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 671 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 612
QY 239 AAACAATCTCAGTGCAGAAAGTAGCTACGATACATCTGGGAATCAACAAG 298
  : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 611 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 552
QY 299 TTACGGTTCATAAAGTGAAGCGGAACAAGTAGTGTTCATTAATAAAGCAGAGATA 358
  : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 551 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 492
QY 359 TGCTACCAGAGATACGACACATGACGATGTTTAAATATTAACAATGAAAAAGTT 418
  : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 491 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 432
QY 419 ATGATCGAAAGATATTACATATAAGATCAGATCAAGGTGACAGCACTGATTTAA 478
  : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 431 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 372
QY 479 GCACATTAACATTAATGTGACAGTAGACATAGCATTTATATAGTGAACAAGTCAA 538
  : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 371 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 312
QY 539 TTACTGATTTGAAAAAGCCTTCCAGGTTCTAAAAATACGTGATATACGAGACA 598
  : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 311 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 252
QY 599 CAATGTATGTAACAATTCACAAGGCGATGCGTATATATAGTTTTTCAATTAAGTACA 658
  : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 251 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 192
QY 659 AAACAAAATTTACGAATGAACAGCAAAAGAGTTGTTAATAATCAGAGCTGTATC 718
  : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 132
QY 719 AAGACATGTTAAGGAAGAGTCAACGGGAATCATTTAATCACTGTCACAAATTA 778
  : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 131 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 72
QY 779 ATGCTATGCCGTATGTGAAGTACTGTATAAAGTGAATTTTAAACAGAGATA 838
  : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 71 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 12
QY 839 AAGATA 844
  : : : : :
Db 11 WWWWWW 6

RESULT 7
AAF58257/c
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1954.
XX
KM Electron-transfer group; ETW; mismatch; genotyping;
KM gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
PI
PI -Umek RM;
XX
WPI; 2001-159728/16.
XX

```

PT Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface

PS Example 6; Page 127; 159pp; English.

CC The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.

SO Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 20.0%; Score 170; DB 22; Length 936;
 Best Local Similarity 1.0%; Pred. No. 2,1e-28;

Matches 8; Conservative 524; Mismatches 254; Indels 0; Gaps 0;

QY 59 GCGGTACAGTAAGATGAGGTTATAGTAAACAGTACCATTACTGTTAAAGTGAGAC 118
 DB 791 GGGCGGCTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178
 QY 119 AGGTGGGTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178
 DB 731 GGGCGGCTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178
 QY * 179 AAAAATTAAGTATGTTGGGATTTGCAAGATTTGAAGTACAAGAAATTTAAACGC 238
 DB 671 GGGCGGCTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178
 QY 239 AAACAATTAAGTATGTTGGGATTTGCAAGATTTGAAGTACAAGAAATTTAAACGC 238
 DB 611 GGGCGGCTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178
 QY 299 TTACGGTTCATTAAGTGAAGCGGGAACAAGTGTCTTTCTATTATTAACGGGAGATA 358
 DB 551 GGGCGGCTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178
 QY 359 TGCTACAGAGATGATGACATGATGATGTTTAAATATTAACAATAAGTAAAGT 418
 DB 491 GGGCGGCTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178
 QY 419 AGTATGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 478
 DB 431 GGGCGGCTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178
 QY 479 GCACATTAACATTAAGTGAAGCGGGAACAAGTGTCTTTCTATTATTAACGGGAGATA 358
 DB 371 GGGCGGCTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178
 QY 539 TTACTGATTTGAAAAAGCTTTCAGGTTCTAAATTAAGTGAATACGAGAGACA 598
 DB 311 GGGCGGCTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178
 QY 599 CAATTATGTAACATTCACAGAGCTATGGGATGATTAAGTGAATACGAGAGACA 658
 DB 251 GGGCGGCTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178
 QY 659 AAACAATTAAGTGAAGCGGGAACAAGTGTCTTTCTATTATTAACGGGAGATA 358
 DB 191 GGGCGGCTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178
 QY 719 AAGAGCATGTAAGAGAGAGTGAACGGAATCATTTAATCATCTGTCACAAATATTA 778
 DB 131 GGGCGGCTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178
 QY 779 ATGCTAATCCGGTATTTGAAGTACTGTAAAGTGAATTAAGTGAATTAAGTGAAT 838
 DB 71 GGGCGGCTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178
 QY 839 AAGATA 844

DB 11 WWWWWWW 6

RESULT 8

AA58259/C

AA58259 standard; DNA; 936 BP.

24-APR-2001 (first entry)

Oligonucleotide D2004.

Electron-transfer group; ETM; mismatch; genotyping;

gene expression; ss.

Synthetic.

MO200107665-A2.

26-JUL-2000; 2000MO-US20476.

26-JUL-1999; 99US-0145695.

17-MAR-2000; 2000US-0190259.

(CLIN-) CLINICAL MICRO SENSORS INC.

Umek RM;

WPI; 2001-159728/16.

Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface

PS Example 6; Page 128; 159pp; English.

CC The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.

SO Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 20.0%; Score 170; DB 22; Length 936;
 Best Local Similarity 1.0%; Pred. No. 2,1e-28;

Matches 8; Conservative 524; Mismatches 254; Indels 0; Gaps 0;

QY 59 GCGGTACAGTAAGATGAGGTTATAGTAAACAGTACCATTACTGTTAAAGTGAGAC 118
 DB 791 GGGCGGCTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178
 QY 119 AGGTGGGTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178
 DB 731 GGGCGGCTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178
 QY 179 AAAAATTAAGTATGTTGGGATTTGCAAGATTTGAAGTACAAGAAATTTAAACGC 238
 DB 671 GGGCGGCTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178
 QY 239 AAACAATTAAGTGAAGCGGGAACAAGTGTCTTTCTATTATTAACGGGAGATA 358
 DB 611 GGGCGGCTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178
 QY 299 TTACGGTTCATTAAGTGAAGCGGGAACAAGTGTCTTTCTATTATTAACGGGAGATA 358
 DB 551 GGGCGGCTCAAGAGTATTTACACGAGCGTGAACAATTACATCAATGATTAAGTAG 178

```
QY 359 TGCCTACGAGATACGACATGTACGATGCTTTTAAATATTAACAATGAAAAAGTT 418
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 491 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 432
QY 419 ATGTATCGAAGATATTACTATAAAGATTCAGATTCAGAGTGAAGCGATTAGATTAA 478
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 431 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 372
QY 479 GCACATTAACATTAATGTGACAGGTACACATAGCAATTATTATAGTGACAAGTCA 538
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 371 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 312
QY 539 TTACTGATTTTGAAGAGCCTTTCCAGGTTCTAAATAAAGTGTGATAACGAAGACA 598
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 311 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 252
QY 599 CAATTGATGTACAAATTCACAAAGGCTATGGTCATATAATAGTTTCAATTAAC 658
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 251 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 192
QY 659 AAACCAAAATTAAGATGAACAGCAAAAGAGTTTGTATAATTCACAAGCTGTATC 718
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 132
QY 719 AAGACATGTTAGAGAGAGTGAACGGGAATCATTTATCATCTGTGCACATATTA 778
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 131 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 72
QY 779 ATGCTAATGCCGTATGTGAAGTACTGTAAAGTGAATTAAGTTTAAAGAGATA 838
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 71 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 12
QY 839 AAGATA 844
   : : : :
Db - 11 WWWWMMW 6

RESULT 9
AAFS8262/c
ID AAF58262 standard; DNA; 936 BP.
XX
AC AAF58262;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2007.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
   gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
   17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI: 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
   hybridization assays, e.g. for genotyping, allowing repeat analyses on
   a single surface
XX
PS Example 6; Page 128; 159pp; English.
```

```
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

Query Match      20.0%; Score 170; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 2,1e-28;
Matches 8; Conservative 524; Mismatches 254; Indels 0; Gaps 0;

QY 59 GCGGTACAGTAAGATAGAGGGTATTAGTAAACAGTACCATTAAGTGAAGC 118
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 791 GGGCWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 732
QY 119 AGGTGGTCAAGCAGTTATTTACACAGACGGTGCAACAATTACATTCATTAAGTAG 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 731 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 672
QY 179 AAAAAATTAAGTATGTTTGGGATTTGCAGAATTGAAATTAAGAAAGAAATTTAACCC 238
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 671 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 612
QY 239 AAACCAATACTCTCAGATGACAAGTAGTACGATAAATCTGGATAAATCAACGATG 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 611 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 552
QY 299 TTACGGTTCATAAAGTGAAGCGGGAACAAGTAGTGTCTTTCTATTATAAAGCGAGATA 358
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 551 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 492
QY 359 TGCCTACGAGATACGACATGTACGATGGTTTAAATTAACAATGAAAAAGTT 418
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 491 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 432
QY 419 ATGTATCGAAGATATTACTATAAAGATCAGATTCAGGTGAGACAGATTAGATTAA 478
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 431 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 372
QY 479 GCACATTAACATTAATGTGACAGGTACACATAGCAATTATTATAGTGACAAGTCA 538
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 371 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 312
QY 539 TTACTGATTTTGAAGAGAGTGAACGGGAATCATTTATCATCTGTGCACATATTA 778
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 311 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 252
QY 599 CAATTGATGTACAAATTCACAAAGGCTATGGTCATATAATAGTTTCAATTAAC 658
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 251 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 192
QY 659 AAACCAAAATTAAGATGAACAGCAAAAGAGTTTGTATAATTCACAAGCTGTATC 718
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 132
QY 719 AAGACATGTTAGAGAGAGTGAACGGGAATCATTTATCATCTGTGCACATATTA 778
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 131 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 72
QY 779 ATGCTAATGCCGTATGTGAAGTACTGTAAAGTGAATTAAGTTTAAAGAGATA 838
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 71 WWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 12
QY 839 AAGATA 844
   : : : :
Db 11 WWWWMMW 6

RESULT 10
```


CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.

Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 19.1%; Score 162.4; DB 22; Length 936;
Best Local Similarity 0.5%; Pred. No. 9.7e-27;
Matches 4; Conservative 522; Mismatches 258; Indels 0; Gaps 0;

```
QY 63 TACAGTAAGATGAGGGTTATAGTAAACAGTACCATTAAGTGTAAAGTGAACAGGT 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6  www. .... 65
QY 123 GGGTCAAGCAGTATTATACACGAGCGTGCAACAATTACATTAAGTAAAGTAGAAAA 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66  www. .... 125
QY 183 ATTAAGTGTGTTTCGGGATTTCCAGATTGGAAGTACAGGAAGAAATTAAACGCAAAAC 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 126 www. .... 185
QY 243 AAATACCTCAGATGACAAAGTACGATACATCGGGATAAATCAACGAAATGTAC 302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 186 www. .... 245
QY 303 GGTTCATAAAGTGAAGCGGGAACAAGTACTGTTCTATTATTAACGAGAGATATGCT 362
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 246 www. .... 305
QY 363 ACCAGAAGATACGACATGACGATGTTTAAATATTAAACAATGAAAAAGTATGT 422
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 306 www. .... 365
QY 423 ATCGAAGATATTACTATAAAGATCAGATTCAAGGTGACAGCAGATTAGATTAAACAC 482
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 366 www. .... 425
QY 483 ATTAACATTATGTGACAGGTACATACATTAATTATAGTGACAAAGTCAATTAC 542
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 426 www. .... 485
QY 543 TGTATTTGAAAAAGCCCTTCCAGGTTCTAAATAACTGTTGATTAATACGAAGACACAT 602
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 486 www. .... 545
QY 603 TGATGTAAACATTCACAAAGCTATGCGTCAATTAATAGTTTCAATTAACTAATAAAC 662
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 546 www. .... 605
QY 663 CAAATATTACGAATGAACAGCAAAAGAGTTTGTAAATTACAAAGCTTGATACAGA 722
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 606 www. .... 665
QY 723 GCATGTGAAGGAAGTGAACGGGAATCATTTAATCATCTGTGCACATTTTAATGC 782
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 666 www. .... 725
QY 783 TAATGCCGATTAAGAGTACTGTAAGGTGAATTAAGTTTAAACAGGATAAGA 842
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 726 www. .... 785
QY 843 TACC 846
   : : |
Db 786 wwg 789
```

RESULT 12
AAFS8254
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX

DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KM Electron-transfer group; ETW; mismatch; genotyping;
XX gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PE 26-JUL-2000; 2000MO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX
PR 17-MAR-2000; 2000US-0190259.
XX

PA (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;

DR WPI; 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
XX

PS Example 6; Page 127; 159pp; English.

XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETW) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.

XX Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 19.1%; Score 162.4; DB 22; Length 936;
Best Local Similarity 0.5%; Pred. No. 9.7e-27;
Matches 4; Conservative 522; Mismatches 258; Indels 0; Gaps 0;

```
QY 63 TACAGTAAGATGAGGGTTATAGTAAACAGTACCATTAAGTGTAAAGTGAACAGGT 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6  www. .... 65
QY 123 GGGTCAAGCAGTATTATACACGAGCGTGCAACAATTACATTAAGTAAAGTAGAAAA 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66  www. .... 125
QY 183 ATTAAGTGTGTTTCGGGATTTCCAGATTGGAAGTACAGGAAGAAATTAAACGCAAAAC 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 126 www. .... 185
QY 243 AAATACCTCAGATGACAAAGTACGATACATCGGAATTAATCAACGAAATGTAC 302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 186 www. .... 245
QY 303 GGTTCATAAAGTGAAGCGGGAACAAGTACTGTTTCTATTATTAACGGAAGTATGCT 362
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 246 www. .... 305
QY 363 ACCAGAAGATACGACATGATGATGTTTAAATATTAAACAATGAAAAAGTATGT 422
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 306 www. .... 365
QY 423 ATCGAAGATATTACTATAAAGATCAGATTCAAGGTGACAGCAGTTAGATTAAACAC 482
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 366 www. .... 425
QY 483 ATTAACATTATGTGACAGGTACACATTAATTATAGTGACAAAGTCAATTAC 542
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2002, 21:27:31 ; Search time 90.75 Seconds
(without alignments)
218.786 Million cell updates/sec

Title: US-09-813-820-3

Sequence: 1 GACGATTAATAATGGAATAAT.....AACAGATTAATACCAAG 849

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

Issued Patents.NA:*
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
3: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
5: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
6: /cgn2_6/prodata/2/ina/5A_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	849	100.0	849	4	US-08-856-253-3	Sequence 3, Appl1
2	849	100.0	1500	4	US-08-856-253-5	Sequence 5, Appl1
3	849	100.0	3827	2	US-08-447-031A-1	Sequence 1, Appl1
4	792.2	93.3	4612	2	US-08-447-031A-8	Sequence 8, Appl1
5	441	51.9	441	4	US-08-856-253-1	Sequence 1, Appl1
6	64.2	7.6	7218	1	US-08-232-463-14	Sequence 14, Appl1
7	47.4	5.6	19124	2	US-08-487-826B-13	Sequence 13, Appl1
8	46	5.4	3095	6	5231168-1	Patent No. 5231168
9	42.8	5.0	1341	4	US-08-960-780-51	Sequence 51, Appl1
10	42.8	5.0	1341	4	US-09-073-898-51	Sequence 51, Appl1
11	42.8	5.0	1341	4	US-09-371-913A-2	Sequence 2, Appl1
12	42.8	5.0	4041	1	US-08-471-033-22	Sequence 22, Appl1
13	42.8	5.0	4041	1	US-08-471-044-19	Sequence 19, Appl1
14	42.8	5.0	4041	2	US-08-463-483A-22	Sequence 22, Appl1
15	42.8	5.0	4041	2	US-08-471-046A-22	Sequence 22, Appl1
16	42.8	5.0	4041	2	US-08-470-566B-22	Sequence 22, Appl1
17	42.8	5.0	4041	2	US-08-469-334-22	Sequence 22, Appl1
18	42.8	5.0	4041	3	US-09-300-529-22	Sequence 22, Appl1
19	42.8	5.0	4074	1	US-08-471-033-19	Sequence 19, Appl1
20	42.8	5.0	4074	1	US-08-471-044-19	Sequence 19, Appl1
21	42.8	5.0	4074	2	US-08-463-483A-19	Sequence 19, Appl1
22	42.8	5.0	4074	2	US-08-471-046A-19	Sequence 19, Appl1
23	42.8	5.0	4074	2	US-08-470-566B-19	Sequence 19, Appl1
24	42.8	5.0	4074	2	US-08-469-334-19	Sequence 19, Appl1
25	42.8	5.0	4074	3	US-09-300-529-19	Sequence 19, Appl1
26	42.8	5.0	6049	1	US-08-471-033-1	Sequence 1, Appl1
27	42.8	5.0	6049	2	US-08-471-044-1	Sequence 1, Appl1

28	42.8	5.0	6049	2	US-08-463-483A-1	Sequence 1, Appl1
29	42.8	5.0	6049	2	US-08-471-046A-1	Sequence 1, Appl1
30	42.8	5.0	6049	2	US-08-470-566B-1	Sequence 1, Appl1
31	42.8	5.0	6049	2	US-08-469-334-1	Sequence 1, Appl1
32	42.8	5.0	6049	3	US-09-300-529-1	Sequence 595, App
33	42.4	5.0	658	4	US-08-998-416-595	Sequence 41, Appl
34	41.6	4.9	1689	2	US-07-991-867B-41	Sequence 41, Appl
35	41.6	4.9	1689	2	US-08-544-332-41	Sequence 41, Appl
36	41.6	4.9	8457	1	US-07-991-867B-1	Sequence 1, Appl1
37	41.6	4.9	8457	2	US-08-544-332-1	Sequence 1, Appl1
38	40.8	4.8	615	4	US-08-998-416-186	Sequence 186, App
39	40.8	4.8	5361	4	US-08-973-462-2	Sequence 2, Appl1
40	40.8	4.8	6152	4	US-08-973-462-1	Sequence 1, Appl1
41	40.4	4.8	6768	1	US-08-107-755A-1	Sequence 1, Appl1
42	40.2	4.7	3393	4	US-09-104-324B-1	Sequence 1, Appl1
43	40.2	4.7	3393	4	US-09-162-713-1	Sequence 1, Appl1
44	39.6	4.7	636	4	US-08-998-416-1137	Sequence 1137, Ap
45	39.6	4.7	1511	1	US-07-991-867B-8	Sequence 8, Appl1

ALIGNMENTS

RESULT 1
US-08-856-253-3
Sequence 3, Application US/08856253
Patent No. 6286214
GENERAL INFORMATION:
APPLICANT: Hook, Magnus
APPLICANT: Patti, Joseph M.
APPLICANT: House-Pompeo, Karen
APPLICANT: Stahanam, Narayana
APPLICANT: Symersky, Jindrich
TITLE OF INVENTION: COLLAGEN BINDING PROTEIN COMPOSITIONS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: U.S.
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/856,253
FILING DATE: Concurrently Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/017,678
FILING DATE: 16-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: TAMR:193
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 849 base pairs
TYPE: nucleic acid
TOPOLOGY: linear
US-08-856-253-3
Query Match 100.0%; Score 849; DB 4; Length 849;
Best Local Similarity 100.0%; Pred. No. 4,5e-194;

QY 481 ACATTAAACATTATGTCAGACGATACATATTTAGTGCACAAAGTGCATT 540
 Db 574 ACATTAAACATTATGTCAGACGATACATATTTAGTGCACAAAGTGCATT 633
 QY 541 ACTGATTTTGAAGAAAGCTTCCAGGTTCTTAATTAACGTGTAATACGAGACACA 600
 Db 634 ACTGATTTTGAAGAAAGCTTCCAGGTTCTTAATTAACGTGTAATACGAGACACA 693
 QY 601 ATTGATGTAAACAATTCACAAAGGCTATGGTCAATATACTTTTTCATTAAGTACAAA 660
 Db 694 ATTGATGTAAACAATTCACAAAGGCTATGGTCAATATACTTTTTCATTAAGTACAAA 753
 QY 661 ACCAAATTTACGATGTAACACGACAAAAGAGTTTGTAAATTAATTCACAAAGTGCATCAA 720
 Db 754 ACCAAATTTACGATGTAACACGACAAAAGAGTTTGTAAATTAATTCACAAAGTGCATCAA 813
 QY 721 GAGCATGTAAGAAAGATGACGAGGAATCATTTAATCACTGTCACAAATATTAT 780
 Db 814 GAGCATGTAAGAAAGATGACGAGGAATCATTTAATCACTGTCACAAATATTAT 873
 QY 781 GCTAATGCGGTAATGTAAGGTAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAG 840
 Db 874 GCTAATGCGGTAATGTAAGGTAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAG 933
 QY 841 GATACCAAG 849
 Db 934 GATACCAAG 942

RESULT 3
 US-08-447-031A-1
 : Sequence 1, Application US/0847031A
 : Patent No. 5851794
 : GENERAL INFORMATION:
 : APPLICANT: GUSS, Bengt
 : APPLICANT: HOOK, Magnus
 : APPLICANT: JONSSON, Hans
 : APPLICANT: LINDBERG, Martin
 : APPLICANT: PATRI, Joseph
 : APPLICANT: SIGNAS, Christer
 : APPLICANT: SWITALSKI, Lech
 : TITLE OF INVENTION: A COLLAGEN BINDING PROTEIN AS WELL AS
 : TITLE OF INVENTION: ITS PREPARATION
 : NUMBER OF SEQUENCES: 8
 : CORRESPONDENCE ADDRESSES:
 : ADDRESSEE: Burns, Doane, Swecker & Mathis
 : STREET: P.O. Box 1404
 : CITY: Alexandria
 : STATE: Virginia
 : COUNTRY: United States
 : ZIP: 22313-1404
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : OPERATING SYSTEM: IBM PC compatible
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/447,031A
 : FILING DATE: 22-MAY-1995
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/861,804
 : FILING DATE: 21-AUG-1992
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: WO PCT/SE91/00707
 : FILING DATE: 22-OCT-1991
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: SE 9003374-7
 : FILING DATE: 22-OCT-1990
 : ATTORNEY/AGENT INFORMATION:
 : NAME: McGowan, Malcolm K.
 : REGISTRATION NUMBER: 39,300

REFERENCE/DOCKET NUMBER: 012889-006
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620
 TELEFAX: (703) 836-2021
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3827 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-447-031A-1

Query Match 100.0%; Score 849; DB 2; Length 3827;
 Best Local Similarity 100.0%; Pred. No. 6,6e-194;
 Matches 849; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACGTTAAATGGAATATACAAATGTGACATGATTAAGTGGCATGGCCGACAGC 60
 Db 331 GACGTTAAATGGAATATACAAATGTGACATGATTAAGTGGCATGGCCGACAGC 390
 QY 61 GGTACAGTAAGATGAGGTTATGTAAGACAGTACCTTACTGTTAAAGTGAACAG 120
 Db 391 GGTACAGTAAGATGAGGTTATGTAAGACAGTACCTTACTGTTAAAGTGAACAG 450
 QY 121 GTGGGTCAAGCATTTATTCACACAGACGTCACAAATTTACATTCATTAAGTAGAA 180
 Db 451 GTGGGTCAAGCATTTATTCACACAGACGTCACAAATTTACATTCATTAAGTAGAA 510
 QY 181 AAATTAAGTATGTTTCCGAGTTTGCAGAAATTTGAAGTCAAGAAAGTAAATTAACGCA 240
 Db 511 AAATTAAGTATGTTTCCGAGTTTGCAGAAATTTGAAGTCAAGAAAGTAAATTAACGCA 570
 QY 241 ACAATTAAGTATGTTTCCGAGTTTGCAGAAATTTGAAGTCAAGAAAGTAAATTAACGCA 300
 Db 571 ACAATTAAGTATGTTTCCGAGTTTGCAGAAATTTGAAGTCAAGAAAGTAAATTAACGCA 630
 QY 301 ACAGTTCAATTAAGTATGTTTCCGAGTTTGCAGAAATTTGAAGTCAAGAAAGTAAATTAACGCA 360
 Db 631 ACAGTTCAATTAAGTATGTTTCCGAGTTTGCAGAAATTTGAAGTCAAGAAAGTAAATTAACGCA 690
 QY 361 CTACAGAAATGATGACGACATGTCAGATGTTTAAATTAATTAACAAAGAAAGTAAAT 420
 Db 691 CTACAGAAATGATGACGACATGTCAGATGTTTAAATTAATTAACAAAGAAAGTAAAT 750
 QY 421 GTATGAAAGATTTACTATTAAGATGATGATTCAGAGTGAACAGTATGATTAAGC 480
 Db 751 GTATGAAAGATTTACTATTAAGATGATGATTCAGAGTGAACAGTATGATTAAGC 810
 QY 481 ACATTAAACATTATGTCAGACGATACATATTTAGTGCACAAAGTGCATT 540
 Db 811 ACATTAAACATTATGTCAGACGATACATATTTAGTGCACAAAGTGCATT 870
 QY 541 ACTGATTTTGAAGAAAGCTTCCAGGTTCTTAATTAACGTGTAATACGAGACACA 600
 Db 871 ACTGATTTTGAAGAAAGCTTCCAGGTTCTTAATTAACGTGTAATACGAGACACA 930
 QY 601 ATTGATGTAAACAATTCACAAAGGCTATGGTCAATATACTTTTTCATTAAGTACAAA 660
 Db 931 ATTGATGTAAACAATTCACAAAGGCTATGGTCAATATACTTTTTCATTAAGTACAAA 990
 QY 661 ACCAAATTTACGATGTAACACGACAAAAGAGTTTGTAAATTAATTCACAAAGTGCATCAA 720
 Db 991 ACCAAATTTACGATGTAACACGACAAAAGAGTTTGTAAATTAATTCACAAAGTGCATCAA 1050
 QY 721 GAGCATGTAAGAAAGATGACGAGGAATCATTTAATCACTGTCACAAATATTAT 780
 Db 1051 GAGCATGTAAGAAAGATGACGAGGAATCATTTAATCACTGTCACAAATATTAT 1110
 QY 781 GCTAATGCGGTAATGTAAGGTAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAG 840
 Db 1111 GCTAATGCGGTAATGTAAGGTAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAG 1170

```

Oy      841 GATGCCAAG 849
|||||||
Db      1171 GATGCCAAG 1179

RESULT      4
US-08-447-031A-8
; Sequence 8, Application US/08447031A
; Patent No. 5851794
; GENERAL INFORMATION:
; APPLICANT: GUSS, Bengt
; APPLICANT: HOOK, Magnus
; APPLICANT: JONSSON, Hans
; APPLICANT: LINDBERG, Martin
; APPLICANT: PATTI, Joseph
; APPLICANT: SIGNAS, Christer
; APPLICANT: SWITALSKI, Lech
; TITLE OF INVENTION: A COLLAGEN BINDING PROTEIN AS WELL AS
; TITLE OF INVENTION: ITS PREPARATION
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,031A
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/861,804
; FILING DATE: 21-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/SE91/00707
; FILING DATE: 22-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 900374-7
; FILING DATE: 22-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 012889-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ. ID NO. 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4612 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 931..4485
;
US-08-447-031A-8

Query Match      93.3%; Score 792.2; DB 2; Length 4612;
Best Local Similarity 97.8%; Pred. No. 2.4e-180;
Matches 836; Conservative 0; Mismatches 13; Indels 6; Gaps
1 GACCATAAATAATGAAATAATACAAATAATGGGACATGATTAAAGTGGCGGCGGCAAGC 60
|||||
Db 1111 GACCATAAATAATGAAATAATACAAATAATGGGACATGATTAAAGTGGCGGCGGCAAGC 1170

```

QY	61	GGTCACTAAAGATAGAGGGTTTACTTAAACAGTACCATTAACGTTAAAGGTGACAG	120
Db	1171	GGTCACTAAAGATAGAGGGTTTACTTAAACAGTACCATTAACGTTAAAGGTGACAG	1220
QY	121	GTGGGTCAAGCAGTTATTATACACAGCGGTGACACATTTACATGATTAAGGTACAA	180
Db	1231	GTGGGTCAAGCAGTTATTATACACAGCGGTGACACATTTACATGATTAAGGTACAA	1290
QY	181	AAATTAAAGTATGTTTCGGGATTTGCAGAAATTTGAAGTACAAAGAAATTTTAACGAA	240
Db	1291	AAATTAAAGTATGTTTCGGGATTTGCAGAAATTTGAAGTACAAAGAAATTTTAACGAA	1350
QY	241	ACAATATCTTCAGTGTGACAAAGTAGCTAGATTAACCTCGGGAAATTAATCAACGATTT	300
Db	1351	ACAATATCTTCAGTGTGACAAAGTAGCTAGATTAACCTCGGGAAATTAATCAACGATTT	1410
QY	301	A - -CGGTTCATTAAGAGTGAAGCGGGAACAAGTAGTGTTCATTATATATAA - -CGGAG	355
Db	1411	ATCGGTTCATTAAGAGTGAAGCGGGAACAAGTAGTGTTCATTATATATATAA - -CGGAG	1470
QY	356	ATATGCTACC - AGAAGTACGACACATGTACGATGGTTTAAATATTAACATGCAAAAA	414
Db	1471	ATATGCTACCAGAAAGTAGACACATGTACGATGGTTTAAATATTAACATGCAAAAA	1530
QY	415	AGTATGTATCGAAAGATATTACTATTAAGAGATCAAGATTCAGGTGACACAGTATAGT	474
Db	1531	AGTATGTATCGAAAGATATTACTATTAAGAGATCAAGATTCAGGTGACACAGTATAGT	1590
QY	475	TTAAGCACATTAACAATTAAATGTGACAGGTACACATGACATTAATTAATGAGCAAAAGT	534
Db	1591	TTAAGCACATTAACAATTAAATGTGACAGGTACACATGACATTAATTAATGAGCAAAAGT	1650
QY	535	GCAATTCGTGATTTTGAAGAAAGCCTTCCAGGTCCTAAATATACGTTGATTAATACGAG	594
Db	1651	GCAATTCGTGATTTTGAAGAAAGCCTTCCAGGTCCTAAATATACGTTGATTAATACGAG	1710
QY	595	AACCAATGATGTAACAATTCACAAAGCCTATGGTCATATAATAGTTTTCAATTATAC	654
Db	1711	AACCAATGATGTAACAATTCACAAAGCCTATGGTCATATAATAGTTTTCAATTATAC	1770
QY	655	TACAAAACCAAAATTACGAATGACAGCAAAAGAGTTGTTAATTAATTCACACAGCTGG	714
Db	1771	TACAAAACCAAAATTACGAATGACAGCAAAAGAGTTGTTAATTAATTCACACAGCTGG	1830
QY	715	TATCAAGAGATGTTAGAGAAAGTAAGCAAGGGAATCATTAATCATACGTGCACAT	774
Db	1831	TATCAAGAGATGTTAGAGAAAGTAAGCAAGGGAATCATTAATCATACGTGCACAT	1890
QY	775	ATTAAATCTAATGCCGTTATTGAAGTACTGTAAAGTGAAATTAAGTTTAAACAG	834
Db	1891	ATTAAATCTAATGCCGTTATTGAAGTACTGTAAAGTGAAATTAAGTTTAAACAG	1950
QY	835	GATTAAGATACCAAG	849
Db	1951	GATTAAGATACCAAG	1965
RESULT 5			
US-08-856-253-1			
; Sequence 1, Application US/08856253			
; Patent No. 6288214			
; GENERAL INFORMATION:			
; APPLICANT: Hook, Magnus			
; APPLICANT: Patti, Joseph M.			
; APPLICANT: House-Pompeo, Karen			
; APPLICANT: Strahan, Narayana			
; APPLICANT: Symeisky, Jindrich			
; TITLE OF INVENTION: COLLAGEN BINDING PROTEIN COMPOSITIONS			
; NUMBER OF SEQUENCES: 8			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Arnold, White & Durkee			

Db 1785 agtagaggttgaagaatcttcacgaattgttgaattgaagaagtaccacaaac 1844
Qy 576 AAGTGTGATATGACGAAGACACATGATGATACATTCACAGGCTTGGTCA 635
Db 1845 aaatacaatgaataatctgaataccttaaacccgaagaagaagaatgaattagt 1904
Qy 636 TAATAGTTTTCAATTAATCAACAAACCAAAATTTACGAATGACGCAAAAGAGTTGT 695
Db 1905 tgaagaaagaacatccacagaagaccggtgtacctacatcaatgaataatgaacgt 1964
Qy 696 TAATTAATTCACAGCTGTGATCAAGAGCATGTGAAGAGAGAGTGAACGGGAATCAT 755
Db 1965 tactcccaacatctgtaggtgaggtccactaacccagatagtagtccaatgaatagt 2024
Qy 756 TAATCATCTGTGCACATATTAATGCTAATGCCGGTATTGAAGTACTGTAAGAGTGA 815
Db 2025 acaagaataataacccaataaagaagaaacacagtagtagtgcctcaaaacatgtag 2084
Qy 816 ATTAAGTTTAAACAGCATTAAGAT 843
Db 2085 acaaatatacagaagatgataatgat 2112

RESULT 9
US-08-960-780-51
; Sequence 51, Application US/08960780
; Patent No. 6204435
; GENERAL INFORMATION:
; APPLICANT: Felleison, Jerald S.
; APPLICANT: Schepf, H. Ernest
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Stockhoff, Brian A.
; APPLICANT: Schmeits, James
; APPLICANT: Loewer, David
; APPLICANT: Dullum, Charles Joseph
; APPLICANT: Muller-Cohn, Judy
; APPLICANT: Stamp, Lisa
; TITLE OF INVENTION: No. 6204435el Pesticidal Toxins and Nucleotide
; TITLE OF INVENTION: Sequences Which Encode These Toxins
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/960,780
; FILING DATE: 30-OCT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,848
; FILING DATE: 30-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA-708
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 177C8
; US-08-960-780-51

Query Match 5.0%; Score 42.8; DB 4; Length 1341;
Best Local Similarity 47.4%; Pred. No. 0.091;
Matches 128; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

Qy 393 TTTAAATATTAACATGAAAAAGTTATGATCGAAGATATTACTATTAAGATCAGAT 452
Db 288 TATAAGACAATTAATAAGAAATTAATCTTTCTATGCGACGCTCATTTGAAGATGAAT 347
Qy 453 TCAAGCTGCACAGCAGTTAGATTTAAGACATTAACATTATATGACAGGTACATAG 512
Db 348 AAAAGATTTAAAGAAATTTAGATTAAGATTTGATTAACCAATCATCATTAATAT 407
Qy 513 CAATTATTAATGATGCAACAGTCGATTAATTTGAAAAAGCCTTCCAGTTCTAA 572
Db 408 CACCTATATAAATGTGCAACCGACACAAATGGATTTAATTAATCTTTAACAGAAAGTAA 467
Qy 573 AATACTGTGATATATACGAAGAACACAAATTGATGAACAAATTCACAAAGGCTATGGTC 632
Db 468 TACGATTAATTCGTGATGCAATGGCACAGCTTAAGAACAATTTTAGATAGGATATTAA 527
Qy 633 ATATTAATGTTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 662
Db 528 GTTGATGATTAATCTAGATGACGATTTAAC 557

RESULT 10
US-09-073-898-51
; Sequence 51, Application US/09073898
; Patent No. 6242669
; GENERAL INFORMATION:
; APPLICANT: Felleison, Jerald S.
; APPLICANT: Schepf, H. Ernest
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Stockhoff, Brian A.
; APPLICANT: Schmeits, James
; APPLICANT: Loewer, David
; APPLICANT: Dullum, Charles Joseph
; APPLICANT: Muller-Cohn, Judy
; APPLICANT: Stamp, Lisa
; APPLICANT: Morrill, George
; APPLICANT: Finstad-Lee, Stacey
; TITLE OF INVENTION: No. 6242669el Pesticidal Toxins and Nucleotide
; TITLE OF INVENTION: Sequences Which Encode These Toxins
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,898
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,848
; FILING DATE: 30-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sanders, Jay M.

REGISTRATION NUMBER: 39,355
 REFERENCE/DOCKET NUMBER: MA-708C1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 352-375-8100
 TELEFAX: 352-372-5800
 INFORMATION FOR SEQ ID NO: 51:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1341 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 INDIVIDUAL ISOLATE: PS177C8
 US-09-073-898-51

Query Match	5.0%	Score 42.8	DB 4	Length 1341
Best Local Similarity	47.4%	Pred. No. 0.091		
Matches 128; Conservative	0	Mismatches 142	Indels 0	Gaps 0

QY	393	TTTAAATTACCACTAAAAAGTTATGATCGAAGATATTACTCTAAAGGATCAGAT	452
Db	288	TATTAACACAATATTAAAGAATTACTTTTCATGCGAGCCATTTGAAGTTGAAT	347
QY	453	TCAAGGTGGACAGCAGTTAGATTAAAGCATTAAACATTATGTCAGAGTACATAG	512
Db	348	AAAAGATTTAAAGCAAAATGTAAGATGTTGATTAACCAACTATCAAAATTCATTAT	407
QY	513	CAATTATTATAGTGGACAAGTGCATTTAGATTGTAAGAAAAGCCCTTCAGGTTCTAA	572
Db	408	CACCTATTAATAATGTGGACCGACACAAATTTGATTTAATTAATCTTTAACGAAAGTAA	467
QY	573	AATTAAGTGTGATTAATACGAAGAACACAATTTGATGTAACTTCCACAGGATGGTC	632
Db	468	TACATTAAATCTGATGCAAGGACAGCTTTAAAGAACAAATTTTACATGAGGATTTAA	527
QY	633	AATATAAGTTTTCATCAATTAACATACTCAAAAC	662
Db	528	GTTTGATAGTATCTAGATAGCGCAATTTAAC	557

```

RESULT 11
US-09-371-913A-2
; Sequence 2, Application US/09371913A
; Patent No. 6297369
; GENERAL INFORMATION:
; APPLICANT: Schaeff, H. Ernest
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Stockhoff, Brian A.
; APPLICANT: Flinstad Lee, Stacey
; APPLICANT: Walz, Mikki
; APPLICANT: Sturgis, Blake
; TITLE OF INVENTION: Pesticidal Toxins and Genes from Bacillus laterosporus
; TITLE OF INVENTION: Strains
; FILE REFERENCE: MA-719XC2 US
; CURRENT APPLICATION NUMBER: US/09/371,913A
; CURRENT FILING DATE: 1999-08-10
; PRIOR APPLICATION NUMBER: 60/095,955
; PRIOR FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: 60/138,251
; PRIOR FILING DATE: 1999-06-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1341
; TYPE: DNA
; ORGANISM: Bacillus laterosporus
; US-09-371-913A-2

```

Query Match	5.0%;	Score 42.8;	DB 4;	Length 1341;
Best Local Similarity	47.4%;	Pred. No. 0.091;		

	Matches	128;	Conservative	0;	Mismatches	142;	Indels	0;	Gaps	0;
QY	393	TTTAAATTTTAAACATGAAAAAAGTATGATCGAAACATATATTACTATATTAAGATCAAGT	452							
Db	288	tataagaacaattataagaattacttctctctgcaagctcatctatgaaagaat	347							
QY	453	TCAAGGTGACAGCAGTTAGATTTAAGCACATTTAAACATTTAATGTGACAGGTACATAG	512							
Db	348	aaagaattaaagaatttgatagatgtctgataaaaccaatctcataatctctat	407							
QY	513	CAATTATATAGTGACACAAAGTGCATTTACGATTTTGA AAAAGCTTCCAGTTTAA	572							
Db	408	cacctataaaatgaggaaaccgacaacaatcggattatataaactttaaaccgaagtta	467							
QY	573	AATAACTGTGATATATAGGAAGAACACAAATTTGATGTAAACAATTCACAAGGCTATGGGTC	632							
Db	468	tagattatattcctgctgaatcgcacagtttaagaacaattttagatagagatataa	527							
QY	633	ATTATAATGTTTTTCAATTAACTACAAAAC	662							
Db	528	gttgatagttatctagatacgcatttaac	557							

RESULT 12
 US-08-471-033-22
 : Sequence 22, Application US/08471033
 : Patent No. 5770696
 : GENERAL INFORMATION:
 : APPLICANT: Warren, Gregory W
 : APPLICANT: Kozietl, Michael G
 : APPLICANT: Mullins, Martha A
 : APPLICANT: Nye, Gordon J
 : APPLICANT: Carr, Brian
 : APPLICANT: Desai, Nalini M
 : APPLICANT: Kostichka, N. Kristy
 : APPLICANT: Duck, Nicholas B
 : APPLICANT: Estruch, Juan J
 : TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
 : NUMBER OF SEQUENCES: 50
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: CIBA-GEIGY Corporation
 : STREET: 7 Skyline Drive
 : CITY: Hawthorne
 : STATE: NY
 : COUNTRY: USA
 : ZIP: 10532
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent In Release #1.0, Version #1.30B
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/471,033
 : FILING DATE:
 : CLASSIFICATION: 530
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/314,594
 : FILING DATE: 09-SEP-1994
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/218,018
 : FILING DATE: 23-MAR-1994
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/037,057
 : FILING DATE: 25-MAR-1993
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Pace, Gary M.
 : REGISTRATION NUMBER: P-40,403
 : REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 919-541-8582
 : TELEFAX: 919-541-8689
 : INFORMATION FOR SEQ ID NO: 22:
 : SEQUENCE CHARACTERISTICS:

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

CONFERENCE ADDRESS:

THIS PAGE BLANK (USPTO)